NON-HOMOGENEOUS LOCAL $T1$ THEOREM: DUAL EXPONENTS

MICHAEL T. LACEY AND ANTTI V. VÄHÄKANGAS

Abstract. We provide an alternative proof of a (local) $T1$ theorem for dual exponents in the non-homogeneous setting of upper doubling measures. This previously known theorem provides necessary and sufficient conditions for the L^p-boundedness of Calderón–Zygmund operators in the described setting, and the novelty lies in the method of proof.

Contents

2000 *Mathematics Subject Classification.* Primary: 42B20 Secondary: 42B25, 60G46.

Research supported in part by grant NSF-DMS 0968499, and a grant from the Simons Foundation (#229596 to Michael Lacey). ML and AVV benefited from the research program Operator Related Function Theory and Time-Frequency Analysis at the Centre for Advanced Study at the Norwegian Academy of Science and Letters in Oslo during 2012–2013. AVV was supported by the Finnish Academy of Science and Letters, Vilho, Yrjö and Kalle Väisälä Foundation.

1. Introduction

1.1. **Background and motivation.** The subject of local Tb theorems in the classical setting of $\mathbf{R}^n$ with Lebesgue measure is rather well understood by now. We refer, in particular, to [17] and to [1–3, 8, 22, 23]. These theorems extend the David–Journé T1 Theorem [7], and the Tb theorem of Christ [6] by giving flexible conditions under which an operator T with a Calderón–Zygmund kernel extends to a bounded linear operator on L^2. By 'local' we understand that the Tb conditions involve a family of test functions b_Q, one for each cube Q, which should satisfy a non-degeneracy condition on its 'own' Q. Furthermore, both b_Q and Tb_Q are subject to normalized integrability conditions on Q (with suitable exponents). Symmetric assumptions are imposed on T^*.

In the non-homogeneous setting less is know. In the relevant literature [16, 19, 26] one usually encounters stronger $L^\infty(\mathbf{R}^n)$ (sometimes BMO) conditions on Tb_Q's, as well as on test functions b_Q. In the search after relaxation of these conditions one faces complications that arise from the feature that the underlying measure μ need not be doubling.

We provide an alternative proof of a local T1 theorem—which is, in fact, a T1 theorem in its local formulation—in the non-homogeneous setting of upper doubling measures, [13, 15]. The local testing functions are indicators of cubes: $b_Q = \mathbf{1}_Q$, and integrability conditions on $\mathbf{1}_Q T \mathbf{1}_Q$ and $\mathbf{1}_Q T^* \mathbf{1}_Q$ are those of dual exponents $1 < p_1 < \infty$ and $p_2 = p_1/(p_1 - 1)$. This result is already known and available in the literature, see Remark 1.4, and the motivation stems from the fact that our novel proof possibly lends itself to other situations. In particular, a non-homogeneous local Tb theorem, say, for dual exponents, has not yet been established, and it seems plausible that the new techniques in the present paper can be used to attack this open and difficult problem.

More precisely, our proof relies upon a so called corona decomposition, adapted to the maximal averages of given two functions f_1 and f_2. The advantage of this approach is that one has powerful quasi orthogonality inequalities, useful throughout the proof. A direct argument can be used to control a difficult 'inside' term, thereby we avoid the typical use of paraproducts and Carleson measures. This argument can be viewed as an extension of its 'homogeneous' counterparts that are developed in [22, 23].

1.2. **A local** T1 **theorem.** Let μ be a compactly supported Borel measure on $\mathbf{R}^n$. We assume the upper doubling conditions of Hytönen [13]: there is a dominating function $\lambda : \mathbf{R}^n \times \mathbf{R}_+ \to \mathbf{R}_+$, and a constant $C_\lambda > 0$, such that for all $x \in \mathbf{R}^n$ and $r > 0$:

$$\mu(B(x, r)) \le \lambda(x, r) \le C_\lambda \lambda(x, r/2)\,.$$

Moreover, we assume that $r \mapsto \lambda(x, r)$ is non-decreasing for all $x \in \mathbf{R}^n$. The number $d = \log_2 C_\lambda$ can be thought of as the dimension of μ.

We assume that a linear operator T is bounded on $L^2(d\mu)$, and it is adapted to λ in the following sense. There is a kernel $K : \mathbf{R}^n \times \mathbf{R}^n \to \mathbf{R}$ such that for all compactly supported $f \in L^2(\mathbf{R}^n)$,

$$Tf(x) = \int_{\mathbf{R}^n} K(x, y) f(y)\, d\mu(y)\,, \qquad x \notin \operatorname{supp}(f).$$

We assume that these kernel estimates hold for some $\eta \in (0, 1)$:

$$|K(x, y)| \leq \min\left\{\frac{1}{\lambda(x, |x-y|)}, \frac{1}{\lambda(y, |x-y|)}\right\}, \qquad x \neq y\,,$$

$$|K(x, y) - K(x', y)| \leq \frac{|x-x'|^\eta}{|x-y|^\eta \lambda(x, |x-y|)}, \qquad |x-y| \geq 2|x-x'|\,, \tag{1.1}$$

and

$$|K(x, y) - K(x, y')| \leq \frac{|y-y'|^\eta}{|x-y|^\eta \lambda(y, |x-y|)}, \qquad |x-y| \geq 2|y-y'|\,.$$

The operator T is said to be a *Calderón–Zygmund operator*. We are interested in *quantitative* estimates for the operator norm of T on $L^p(\mu)$ for $1 < p < \infty$, and the following hypothesis, together with kernel assumptions, provides the essential quantitative information.

- *Local Testing Condition Hypothesis*. For given two exponents $p_1, p_2 \in (1, \infty)$, there is a constant $\mathbf{T}_{\mathrm{loc}}$ as follows. For all cubes Q in $\mathbf{R}^n$,

$$\int_Q |T\mathbf{1}_Q|^{p_1}\, d\mu(x) \leq \mathbf{T}_{\mathbf{loc}}^{p_1} \mu(Q)\,, \qquad \int_Q |T^*\mathbf{1}_Q|^{p_2}\, d\mu(x) \leq \mathbf{T}_{\mathrm{loc}}^{p_2} \mu(Q)\,. \tag{1.2}$$

We provide a novel proof of the following previously known theorem.

1.3. **Theorem.** *Let* T *be a Calderón–Zygmund operator. Fix* $1 < p_1, p_2 < \infty$, $1/p_1 + 1/p_2 \leq 1$. *Assume the following two conditions (1)–(2):*

(1) T *is (a priori) bounded on* $L^{p_1}(d\mu)$*;*
(2) T *satisfies a Local Testing Condition Hypothesis with exponents* p_1 *and* p_2.

Under these assumptions, we have a quantitative norm estimate

$$\mathbf{T} := \|T\|_{L^{p_1}(d\mu) \to L^{p_1}(d\mu)} \lesssim 1 + \mathbf{T}_{\mathrm{loc}}\,,$$

where the implied constant depends on n, p_1, p_2, η, μ.

In the sequel, unless otherwise specified, we assume that p_1 and p_2 are in duality: $p_2 = \frac{p_1}{p_1 - 1}$.

1.4. *Remark.* Theorem 1.3 is known and available in the literature. Indeed, under the assumptions of this theorem, it is straightforward to verify that T satisfies a 'weak boundedness property' and 'testing conditions', namely for all cubes Q in $\mathbf{R}^n$, and an appropriate $\sigma \geq 1$,

$$\left| \int_Q T\mathbf{1}_Q \, d\mu \right| \leq \mathbf{T}_{\mathsf{loc}}\mu(Q) \,, \qquad T\mathbf{1} \in \mathrm{BMO}_\sigma^{p_1}(\mu), \qquad T^*\mathbf{1} \in \mathrm{BMO}_\sigma^{p_2}(\mu)\,; \tag{1.5}$$

we refer to Remark 2.8 for further details. It remains to apply a non-homogeneous T1 theorem, see [25] or [11, Tb theorem 2] for $\lambda(x,r) = r^d$ dominating the measure, and [24, Theorem 2.1] for the general case. Moreover, by using the last theorem, it is even possible to relax the integrability conditions in (1.2) to exponents $p_1 = 1 = p_2$. Let us also remark that the case of $p_1 = 2 = p_2$ has been addressed in [28] with a function $\lambda(x,r) = \max\{\delta(x)^d, r^d\}$ dominating the measure, where $\delta(x) = \mathrm{dist}(x, \mathbf{R}^n \setminus H)$ for an an open set H in $\mathbf{R}^n$.

1.6. *Remark.* The p-independence property of Calderón–Zygmund operators, i.e., if their L^2 boundedness is equivalent to their L^p boundedness, has been addressed, for instance, in [9, 14]. It is an interesting question, if our proof can be adapted to obtain a quantitative p-independence result for Calderón–Zygmund operators, under an appropriate set of local testing hypotheses.

1.3. **Structure of the paper.** We use the non-homogeneous techniques of [25], in particular, good and bad cubes are applied in a partially novel manner. Martingale techniques, including L^p estimates for martingale transforms and Stein's ineguality, are fundamental. These techniques are also applied in a related paper [19], from which we borrow also some other ideas, e.g., treatments of 'separated' and 'nearby' terms. Our main technical contribution is treatment of the most difficult 'inside' term by a strong definition of goodness and a corona decomposition, avoiding (a) explicit construction of paraproduct operators; and (b) Carleson embedding theorems.

The heart of the matter is estimation of a form $|\langle Tf_1, f_2\rangle|$, where f_j's are perturbed functions, supported on large dyadic cubes $Q_{j,0} \in \mathcal{D}_j$. Here $\mathcal{D}_j$ is a random dyadic system. The perturbation is simply a projection to good cubes, and results in that the usual martingale differences $\Delta_Q f_j$ vanish if $Q \subset Q_{j,0}$ is a bad. After a probabilistic absorption argument, the focus will be on a triangular form

$$\Big| \sum_{P,Q \text{ good}} \mathbf{1}_{\ell Q \leq \ell P} \cdot \langle T\Delta_P f_1, \Delta_Q f_2 \rangle \Big| \,,$$

where always $P \subset Q_{1,0}$ and $Q \subset Q_{2,0}$. This form is further split into 'inside', 'separated', and 'nearby' terms. The analysis of the inside term, in which Q is deeply inside P, is taken up in sections 5 and 6—the argument is transparent, and our strong definition of *goodness of cubes* has

a key role. The construction of paraproducts is avoided, and even Carleson embedding theorems are not needed; in this we follow [20, 21]. We apply a corona decomposition, and the associated stopping tree is constructed in Section 4, where we also record the basic 'quasi-orthogonality' properties. The separated term, in which Q is always far away from P, is analysed in Section 7, and the (usual) goodness is crucial. Throughout sections 8–10, we treat the nearby term, where cubes are close to each other both in position and size. The usual surgery is performed.

Acknowledgment. The authors would like to thank Tuomas Hytönen and Henri Martikainen for indicating the connection of Theorem 1.3 to the T1 theorems that are available in the literature.

2. Preliminaries

2.1. **Notation.** The implied constants are allowed to depend upon parameters $r, n, p_1, p_2, \eta, \mu$. The distances are measured in supremum norm, $|x| = \|x\|_\infty$ for $x \in \mathbf{R}^n$. We denote $L^p = L^p(d\mu)$ if $1 \le p \le \infty$. For a cube Q and $f \in L^1_{\mathrm{loc}}$, write $\langle f\rangle_Q := \mu(Q)^{-1}\int_Q f\, d\mu$ with the convention $\langle f\rangle_Q = 0$ if $\mu(Q) = 0$. The side length of a cube Q is written as ℓQ, and the midpoint as x_Q. The 'long distance' between cubes Q and P is $D(Q,P) = \ell Q + \mathrm{dist}(Q,P) + \ell P$.

A 'dyadic cube' is any cube in either random grid $\mathcal{D}_j$ with $j \in \{1,2\}$, Section 2.2. By $\mathcal{D}_{j,k}$ we denote those dyadic cubes $Q \in \mathcal{D}_j$ for which $\ell Q = 2^k$, $k \in \mathbf{Z}$. The dyadic children of $Q \in \mathcal{D}_j$ are $\{Q_1, \ldots, Q_{2^n}\} = \mathrm{ch}(Q)$, its dyadic parent is $\pi_j Q = \pi_j^1 Q$, and $\pi_j^t Q = \pi_j(\pi_j^{t-1} Q)$ for $t \in \{2,3,\ldots\}$. For $\mathcal{S}_j \subset \mathcal{D}_j$ the family $\mathrm{ch}_{\mathcal{S}_j}(S) = \mathrm{ch}^1_{\mathcal{S}_j}(S)$ consists of the $\mathcal{S}_j$-children of $S \in \mathcal{S}_j$: the maximal cubes in $\mathcal{S}_j$ that are strictly contained in S. We also denote $\mathrm{ch}^0_{\mathcal{S}_j}(S) = \{S\}$ and, for $t > 1$, write $S' \in \mathrm{ch}^t_{\mathcal{S}_j}(S)$ if $S' \in \mathrm{ch}_{\mathcal{S}_j}(S'')$ for some $S'' \in \mathrm{ch}^{t-1}_{\mathcal{S}_j}(S)$. For *any* cube Q which is contained in a cube in $\mathcal{S}_j$, we take $\pi_{\mathcal{S}_j} Q = \pi^0_{\mathcal{S}_j} Q$ to be the $\mathcal{S}_j$-parent of Q: the minimal $\mathcal{S}_j$-cube containing Q (if Q is not contained in a cube in $\mathcal{S}_j$, we set $\pi_{\mathcal{S}_j} Q = \mathbf{R}^n$). For $t \ge 1$ and any cube Q, contained in at least $t+1$ cubes in $\mathcal{S}_j$, we let $\pi^t_{\mathcal{S}_j} Q$ to be $\pi^{t-1}_{\mathcal{S}_j} S'$, where $\pi_{\mathcal{S}_j} Q \in \mathrm{ch}_{\mathcal{S}_j}(S')$.

2.2. **Random grids.** We use the foundational tool of random grids, initiated by Nazarov–Treil–Volberg [26], which has in turn been used repeatedly. We refer, e.g., to [10, 15, 21, 27]. Throughout the paper, we shall use two random dyadic grids (systems) $\mathcal{D}_j$, $j \in \{1,2\}$. A third random grid $\mathcal{D}_3$ appears at the very end. These are constructed as follows; we refer to [12] for further details.

The random grids $\mathcal{D}_j$ are parametrized by sequences $\omega_j \in (\{0,1\}^n)^{\mathbf{Z}}$, $j \in \{1,2,3\}$, where we tacitly assume three independent copies of $(\{0,1\}^n)^{\mathbf{Z}}$. More precisely, for a cube $\widehat{Q} \in \widehat{\mathcal{D}}$ in the

standard dyadic grid, the *position* of an ω_j-translated cube is

$$Q = \widehat{Q} \dot{+} \omega_j := \widehat{Q} + \sum_{k:\,2^{-k}<\ell\widehat{Q}} 2^{-k}\omega_{j,k}\,,$$

which is a function of $\omega_j \in (\{0,1\}^n)^{\mathbf{Z}}$. A dyadic grid (system)

$$\mathcal{D}_j = \mathcal{D}(\omega_j) = \{\widehat{Q}\dot{+}\omega_j \,:\, \widehat{Q} \in \widehat{\mathcal{D}}\}$$

is the family of these ω_j-translated cubes. The natural uniform probability measure $\mathbf{P}_{\omega_j}$ is placed upon the respective copy of $(\{0,1\}^n)^{\mathbf{Z}}$. Each component $\omega_{j,k}$, $k \in \mathbf{Z}$, has an equal probability 2^{-n} of taking any of the 2^n values, and all components are independent of each other. The expectation with respect to $\mathbf{P}_{\omega_j}$ is denoted by $\mathbf{E}_{\omega_j}$. We will usually simply write P or S for a cube in $\mathcal{D}_1$, and Q or R for a cube in $\mathcal{D}_2$, instead of the heavier notation $\widehat{Q}\dot{+}\omega_j$ with $\widehat{Q} \in \widehat{\mathcal{D}}$.

Choose, once and for all, a constant $\gamma \in (0,1)$ such that

$$d\gamma/(1-\gamma) \le \eta/4, \quad \gamma \le \frac{\eta}{2(d+\eta)}, \qquad d = \log_2 C_\lambda\,. \tag{2.1}$$

Here η is the constant appearing in the kernel condition (1.1). We also denote

$$\theta(j) = \left\lceil \frac{\gamma j + r}{1-\gamma} \right\rceil \quad \text{for } j = 0,1,2,\ldots.$$

Throughout $r \in \mathbf{N}$ should be thought of as a large integer, whose exact value is assigned later.

2.3. **Goodness of cubes.** We impose a strong definition of goodness: by doing so, we ensure that good cubes $Q \in \mathcal{D}_1 \cup \mathcal{D}_2$ from either system are always far away from the boundaries of much larger cubes in *either* one of these two systems.

A cube $Q \in \mathcal{D}_j$ is k-*bad* for $j,k \in \{1,2\}$ if there is a cube $P \in \mathcal{D}_k$ such that $\ell P \ge 2^r \ell Q$ and $\operatorname{dist}(Q,\partial P) \le (\ell Q)^\gamma (\ell P)^{1-\gamma}$. Otherwise, Q is k-*good*. The following properties are known, [12].

(1) For $\widehat{Q} \in \widehat{\mathcal{D}}$, position and k-goodness of $Q = \widehat{Q}\dot{+}\omega_j$ are independent random variables.
(2) The probability $\pi_{j,k,\text{good}} := \mathbf{P}_{\omega_k}(\widehat{Q}\dot{+}\omega_j \text{ is } k\text{-good})$ is independent of $\widehat{Q} \in \widehat{\mathcal{D}}$.
(3) $\pi_{j,k,\text{bad}} := 1 - \pi_{j,k,\text{good}} \lesssim 2^{-\gamma r}$, with implied constant independent of r.

A cube $Q \in \mathcal{D}_j$ with $j \in \{1,2\}$ is *bad* if it is k-bad for some $k \in \{1,2\}$. Otherwise, we say that Q is *good*. To state this condition otherwise, if $Q \in \mathcal{D}_j$ is good, we have inequality

$$(\ell Q)^\gamma (\ell P)^{1-\gamma} < \operatorname{dist}(Q,\partial P)\,,$$

if $P \in \mathcal{D}_1 \cup \mathcal{D}_2$ and $2^r \ell Q \le \ell P$. Define bad and good projections by $I = P_{j,\text{bad}} + P_{j,\text{good}}$, where

$$P_{j,\text{bad}}\phi := \sum_{Q\in\mathcal{D}_j\,:\,Q \text{ is bad}} \Delta_Q \phi\,, \qquad \phi \in L^q \quad (1<q<\infty)\,.$$

Here $\Delta_Q \phi = \sum_{Q' \in \mathrm{ch}(Q)} \{\langle \phi \rangle_{Q'} - \langle \phi \rangle_Q\} \mathbf{1}_{Q'}$ is the martingale difference with respect to μ. The following proposition is a straightforward modification of [23, Proposition 2.4].

2.2. **Proposition.** *For every $j \in \{1,2\}$ and $1 < q < \infty$ there is a constant $c_q > 0$ so that*

$$\mathbf{E}_{\omega_1}\mathbf{E}_{\omega_2} \|P_{j,\mathrm{bad}}\phi\|_q^q \lesssim 2^{-\gamma r/c_q} \|\phi\|_q^q\,,$$

where $\phi \in L^q$ is any function, independent of both random grids $\mathcal{D}_k$ with $k \in \{1,2\}$. Moreover, the implied constant is independent of r.

Proof. We apply Marcinkiewicz interpolation theorem to the linear operator

$$P_{j,\mathrm{bad}} : L^q(d\mu) \to L^q(\mathbf{P}_{\omega_1} \otimes \mathbf{P}_{\omega_2} \otimes d\mu)\,.$$

The projection to bad cubes is a martingale transform: by inequality (2.3), the following inequality with no decay holds,

$$\mathbf{E}_{\omega_1}\mathbf{E}_{\omega_2} \|P_{j,\mathrm{bad}}\phi\|_p^p \leq \sup_{\omega_1,\omega_2} \|P_{j,\mathrm{bad}}\phi\|_p^p \lesssim \|\phi\|_p^p\,, \quad 1 < p < \infty\,.$$

Thus, it suffices to verify the claimed decay for $q = 2$. To this end, we have by orthogonality of martingale differences,

$$\begin{aligned}
\mathbf{E}_{\omega_1}\mathbf{E}_{\omega_2} \|P_{j,\mathrm{bad}}\phi\|_2^2 &= \mathbf{E}_{\omega_1}\mathbf{E}_{\omega_2} \sum_{\widehat{Q} \in \widehat{\mathcal{D}}} \mathbf{1}_{\widehat{Q}\dot{+}\omega_j \text{ is bad}} \|\Delta_{\widehat{Q}\dot{+}\omega_j}\phi\|_2^2 \\
&\leq \sum_{k=1}^{2} \mathbf{E}_{\omega_1}\mathbf{E}_{\omega_2} \sum_{\widehat{Q} \in \widehat{\mathcal{D}}} \mathbf{1}_{\widehat{Q}\dot{+}\omega_j \text{ is k-bad}} \|\Delta_{\widehat{Q}\dot{+}\omega_j}\phi\|_2^2 \\
&\leq \sum_{k=1}^{2} \pi_{j,k,\mathrm{bad}} \mathbf{E}_{\omega_1}\mathbf{E}_{\omega_2} \sum_{\widehat{Q} \in \widehat{\mathcal{D}}} \|\Delta_{\widehat{Q}\dot{+}\omega_j}\phi\|_2^2 \leq (\pi_{j,1,\mathrm{bad}} + \pi_{j,2,\mathrm{bad}}) \|\phi\|_2^2\,.
\end{aligned}$$

In the third step, we used Fubini's theorem, linearity of expectation, and the fact that $\|\Delta_{\widehat{Q}\dot{+}\omega_j}\phi\|_2^2$ and k-badness of $\widehat{Q}\dot{+}\omega_j$ are independent random variables. ◻

2.4. **Square function inequalities.** The martingale transform inequality is this, see e.g. [5]. For all functions $f \in L^p$, and constants satisfying $\sup_{Q \in \mathcal{D}_j} |\varepsilon_Q| \leq 1$,

$$\Big\| \sum_{Q \in \mathcal{D}_j} \varepsilon_Q \Delta_Q f \Big\|_p \lesssim ||f||_p, \quad 1 < p < \infty\,, \quad j \in \{1,2\}\,. \tag{2.3}$$

A consequence of Khintchine's inequality and inequality (2.3) is the following.

$$\Big\| \Big(\sum_{k\in\mathbf{Z}} |\Delta_{j,k} f|^2 \Big)^{1/2} \Big\|_p \lesssim \|f\|_p \,, \tag{2.4}$$

where $f \in L^p$ with $1 < p < \infty$, and $\Delta_{j,k} f = \sum_{Q\in\mathcal{D}_{j,k}} \Delta_Q f$ for $k \in \mathbf{Z}$ and $j \in \{1, 2\}$.

We will use the following *Stein's inequality*, see e.g. [4]. For $1 < p < \infty$ and $j \in \{1, 2\}$,

$$\Big\| \Big(\sum_{k\in\mathbf{Z}} |\mathbf{E}_{j,k} f_k|^2 \Big)^{1/2} \Big\|_p \lesssim \Big\| \Big(\sum_{k\in\mathbf{Z}} |f_k|^2 \Big)^{1/2} \Big\|_p \,, \tag{2.5}$$

where $(f_k)_{k\in\mathbf{Z}}$ is *any* sequence in $L^p(d\mu)$, $\mathbf{E}_{j,k} f = \sum_{Q\in\mathcal{D}_{j,k}} \mathbf{E}_Q f$, and $\mathbf{E}_Q f = \langle f\rangle_Q \mathbf{1}_Q$. We don't rely on Fefferman–Stein inequalities for the vector-valued maximal function. Stein's inequality is their replacement in the present, non-homogeneous, setting.

2.5. **Off-diagonal estimates.** Here we collect useful off-diagonal estimates.

2.6. **Lemma.** *Let* $Q \subset P \subset R$ *be cubes in* $\mathbf{R}^n$ *such that* $\ell Q \le \operatorname{dist}(Q, R \setminus P)$. *Then,*

$$|T\mathbf{1}_{R\setminus P}(x) - T\mathbf{1}_{R\setminus P}(x_Q)| \lesssim \Big(\frac{\ell Q}{\operatorname{dist}(Q, R\setminus P)} \Big)^{\eta}, \qquad x \in Q \,. \tag{2.7}$$

Proof. The kernel condition (1.1) applies,

$$\mathrm{LHS}(2.7) \le \int_{R\setminus P} |K(x,y) - K(x_Q, y)| \, d\mu(y) \le \int_{R\setminus P} \frac{|x - x_Q|^\eta}{|x-y|^\eta \lambda(x, |x-y|)} \, d\mu(y) \,.$$

Let us denote $\delta := \operatorname{dist}(Q, R\setminus P)$ and $A_j = \{y \in \mathbf{R}^n : 2^j\delta \le |x-y| < 2^{j+1}\delta\}$ for $j \ge 0$. Observe that $R \setminus P \subset \cup_{j=0}^\infty A_j$. Since $A_j \subset B(x, 2|x-y|)$ for each $y \in A_j$, we can bound the last integral by $C_\lambda (\ell Q)^\eta \sum_{j=0}^\infty (2^j\delta)^{-\eta} \lesssim (\ell Q/\delta)^\eta$ as required. □

2.8. *Remark.* Let us verify that the a priori boundedness of T on L^{p_1}, and Local Testing Condition Hypothesis, together imply the assumptions of a T1 theorem; namely, conditions (1.5) with $\sigma = 3$. The first condition therein is, indeed, a trivial consequence of inequality (1.2). Hence, it suffices to verify that $b := T\mathbf{1}$ satisfies $b \in \mathrm{BMO}^{p_1}_\sigma(\mu)$, i.e.,

$$\|b\|_{\mathrm{BMO}^{p_1}_\sigma(\mu)} := \sup_Q \Big\{ \frac{1}{\mu(\sigma Q)} \int_Q |b(x) - \langle b\rangle_Q|^{p_1} \, d\mu(x) \Big\}^{1/p_1} \lesssim 1 + \mathbf{T}_{\mathrm{loc}} \,, \tag{2.9}$$

where the supremum is taken over all cubes Q in $\mathbf{R}^n$. Indeed, a completely analogous argument then shows that $T^*\mathbf{1} \in \mathrm{BMO}^{p_2}_\sigma(\mu)$.

In order to verify inequality (2.9), let us fix a cube Q in which the supremum above is (almost) attained. Let us then fix a large cube R in $\mathbf{R}^n$, containing both $3Q$ and the compact support of the measure μ. In particular, $T\mathbf{1} = T\mathbf{1}_R \in L^{p_1}$, and we can estimate

$$\begin{aligned}\|b\|^{p_1}_{\mathrm{BMO}^{p_1}_\sigma(\mu)} &\lesssim \frac{1}{\mu(3Q)}\int_Q |T\mathbf{1}_R(x) - T\mathbf{1}_{R\setminus 3Q}(x_Q)|^{p_1}\,d\mu(x)\\ &\lesssim \frac{1}{\mu(3Q)}\int_Q |T\mathbf{1}_{3Q}|^{p_1}\,d\mu + \frac{1}{\mu(3Q)}\int_Q |T\mathbf{1}_{R\setminus 3Q} - T_{R\setminus 3Q}(x_Q)|^{p_1}\,d\mu\,.\end{aligned}$$

By inequality (1.2), the first term in the last line is dominated by $\mathbf{T}^{p_1}_{\mathrm{loc}}$. And by Lemma 2.6, the last term is seen to be bounded by $\lesssim 1$.

In the following two lemmata, we write $D(Q,P)/\ell P \sim 2^u$ if $2^u < D(Q,P)/\ell P \le 2^{u+1}$.

2.10. **Lemma.** *Suppose $P \in \mathcal{D}_{1,k}$ and $Q \in \mathcal{D}_{2,k-m}$ is a good cube such that $D(Q,P)/\ell P \sim 2^u$, where $k \in \mathbf{Z}$ and $u, m \in \mathbf{N}_0$. Then, we have $Q \subset \pi_1^{u+\theta(u+m)}P$.*

Proof. Denote $t = u + \theta(u+m) \ge r$. By goodness, either $Q \subset \pi^t P$ or $Q \subset \mathbf{R}^n \setminus \pi^t P$. In the former case, we are done. In the latter case, we obtain a contradiction. Indeed, by goodness,

$$(\ell Q)^\gamma (\ell \pi^t P)^{1-\gamma} < \mathrm{dist}(Q, \partial \pi^t P) = \mathrm{dist}(Q, \pi^t P) \le D(Q,P) \le 2^{u+1}\ell P\,.$$

Substituting $\ell Q = 2^{k-m}$ and $\ell P = 2^k$ yields $u + \theta(u+m) = t < u + \theta(u+m)$ after elementary manipulations. This is a contradiction. □

2.11. **Lemma.** *Suppose that P and Q are as in Lemma 2.10. Assume also that $\ell Q \le \mathrm{dist}(Q,P)$. Then, by denoting $S := \pi_1^{u+\theta(u+m)}P$,*

$$|K(x,y) - K(x_Q,y)| \lesssim \frac{2^{-\eta(u+m)/4}}{\mu(S)}, \qquad (x,y) \in Q \times P\,. \tag{2.12}$$

Proof. By inequalities (1.1) and $\ell Q \le \mathrm{dist}(Q,P)$, we obtain LHS(2.12) $\le \alpha \cdot \beta$ where $\alpha = C^\kappa_\lambda (\ell Q)^\eta / \mathrm{dist}(Q,P)^\eta$ and $\beta = 1/\mu(B(x, 2^\kappa |x-y|))$ with κ specified in the two case studies.

Case $\ell P < \mathrm{dist}(Q,P)$. Choose $\kappa = 2 + \theta(u+m)$. Observe the inequality $2^{u+k} < 4\mathrm{dist}(Q,P)$. Combined with Lemma 2.10 this implies a relation $S \subset B(x, 2^\kappa|x-y|)$ and, in particular, that $\beta \le \mu(S)^{-1}$. The inequality $2^{u+k} < 4\mathrm{dist}(Q,P)$, followed by $C^\kappa_\lambda = 2^{d\kappa}$ and (2.1), shows that $\alpha \lesssim 2^{d\theta(u+m) - \eta(u+m)} \lesssim 2^{-\eta(u+m)/4}$.

Case $\ell P \ge \mathrm{dist}(Q,P)$. Choose $\kappa \in \mathbf{N}$ in such a way that

$$2^{\kappa-1} < \frac{c\,\ell S}{(\ell Q)^\gamma (\ell P)^{1-\gamma}} \le 2^\kappa\,, \qquad c = 2^{r(1-\gamma)}\,.$$

A useful consequence of goodness is this.

$$\operatorname{dist}(Q,P) > (\ell Q)^{\gamma}(\ell P)^{1-\gamma}/c\,. \tag{2.13}$$

Lemma 2.10 and inequality (2.13) yield $S \subset B(x, 2^{\kappa}|x-y|)$, hence $\beta \le \mu(S)^{-1}$. Inequality (2.13) also allows us to estimate

$$\alpha \lesssim 2^{d\kappa}\left(\frac{\ell Q}{\ell P}\right)^{\eta(1-\gamma)} \lesssim 2^{d(m+u+\theta(u+m))-m(d+\eta)(1-\gamma)} \lesssim 2^{-\eta(u+m)/4}\,.$$

In the last step, we used the fact that $u \le 1$ and both of the inequalities (2.1). □

3. Perturbations and a basic decomposition

Let us denote $\mathbf{T} := \|T\|_{L^{p_1}\to L^{p_1}}$. We fix functions $\widetilde{f}_j \in L^{p_j}(d\mu)$, $j = 1,2$, supported in $\operatorname{supp}(\mu)$, and satisfying $\mathbf{T} \le 2|\langle T\widetilde{f}_1, \widetilde{f}_2\rangle|$ and $\|\widetilde{f}_1\|_{p_1} = 1 = \|\widetilde{f}_2\|_{p_2}$.

For almost every pair $\{\mathcal{D}_j : j \in \{1,2\}\}$ we will define certain perturbations $f_j = f_j(\widetilde{f}_j, \mathcal{D}_1, \mathcal{D}_2)$ of functions $\widetilde{f}_j$. The role of these perturbations is indicated by following proposition.

3.1. **Proposition.** *Under assumptions of Theorem 1.3, the following statement holds for a fixed* $t > p_1 \vee p_2$. *For every sufficiently large* $r \in \mathbf{N}$ *and every* $\epsilon, \upsilon \in (0,1)$,

$$\mathbf{E}_{\omega_1}\mathbf{E}_{\omega_2}\|\widetilde{f}_j - f_j\|_{p_j}^{p_j} \le c2^{-\gamma r/c}\,, \qquad j = 1,2\,, \tag{3.2}$$

$$\mathbf{E}_{\omega_1}\mathbf{E}_{\omega_2}\Big|\langle Tf_1, f_2\rangle\Big| \le C(r,\upsilon,\epsilon)(1+\mathbf{T}_{\mathsf{loc}}) + \Big(C(r,\upsilon)\epsilon^{1/t} + C(r)\upsilon^{1/t}\Big)\mathbf{T}\,. \tag{3.3}$$

Aside from the parameters indicated, constants c, $C(r,\upsilon,\epsilon)$, $C(r,\upsilon)$, *and* $C(r)$ *are also allowed to depend upon* n, p_1, p_2, η, μ.

Proposition 3.1 and an absorption argument provide a proof of Theorem 1.3. Hence, we are left with proving this proposition. During this section, we select functions f_j by using projections to good cubes, and then begin with the analysis of the resulting form $\Big|\langle Tf_1, f_2\rangle\Big|$.

3.1. Perturbations of $\widetilde{f}_j$.

For $j \in \{1,2\}$ we denote by $Q_{j,0}$ a cube in $\mathcal{D}_j = \mathcal{D}(\omega_j)$, containing the support $\operatorname{supp}(\mu)$ of the measure μ. Such a cube exists almost surely with respect to ω_j, [19, Lemma 2.8]. In the sequel, we will restrict ourselves to such sequences ω_j. Let $\mathcal{G}_j$ be the family of all good cubes in $\mathcal{D}_j$ that are contained in $Q_{j,0}$, and denote $\mathcal{G}_{j,k} = \mathcal{D}_{j,k} \cap \mathcal{G}_j$ for $k \in \mathbf{Z}$.

We define approximates of the functions $\widetilde{f}_j$ to be the following perturbations,

$$f_j := \langle \widetilde{f}_j\rangle_{Q_{j,0}}\mathbf{1}_{Q_{j,0}} + \sum_{Q\in\mathcal{G}_j} \Delta_Q \widetilde{f}_j\,, \qquad j = 1,2\,.$$

Recall the fact that the support of μ is contained in $Q_{j,0}$. Therefore $\Delta_Q \widetilde{f}_j = 0$ almost everywhere w.r.t. μ if $Q \in \mathcal{D}_j$ is not contained in $Q_{j,0}$. Hence, in the view of Proposition 2.2, we have

$$\mathbf{E}_{\omega_1}\mathbf{E}_{\omega_2}\|\widetilde{f}_j - f_j\|_{p_j}^{p_j} = \mathbf{E}_{\omega_1}\mathbf{E}_{\omega_2}\|P_{j,\mathsf{bad}}\widetilde{f}_j\|_{p_j}^{p_j} \le c2^{-\gamma r/c}\,.$$

This is inequality (3.2).

3.2. **Decomposition of the bilinear form.** During the course of the remaining sections, we prove inequality (3.3), which then completes the proof of Theorem 1.3.

By using the facts that $f_j = f_j \mathbf{1}_{Q_{j,0}}$ and $\Delta_R f_j = 0$ if $R \subset Q_{j,0}$ is a bad cube, we easily find that an expansion of the bilinear form is

$$\langle Tf_1, f_2\rangle = \langle T\mathbf{E}_{Q_{1,0}} f_1, f_2\rangle + \langle T \sum_{P\in\mathcal{G}_1} \Delta_P f_1, \mathbf{E}_{Q_{2,0}} f_2\rangle + \sum_{(P,Q)\in\mathcal{G}_1\times\mathcal{G}_2} \langle T\Delta_P f_1, \Delta_Q f_2\rangle. \tag{3.4}$$

Using the assumptions and inequality (2.3), it is straightforward to verify that

$$|\langle T\mathbf{E}_{Q_{1,0}} f_1, f_2\rangle| + |\langle T \sum_{P\in\mathcal{G}_1} \Delta_P f_1, \mathbf{E}_{Q_{2,0}} f_2\rangle| \lesssim \mathbf{T}_{\mathsf{loc}} \|f_1\|_{p_1}\|f_2\|_{p_2} \lesssim \mathbf{T}_{\mathsf{loc}}\,.$$

The last term in the right hand side of (3.4) remains. This main term is further split into dual triangular sums, one of which is the sum over $(P,Q) \in \mathcal{G}_1 \times \mathcal{G}_2$ such that $\ell P \ge \ell Q$. This sum will be our main point of interest, and we only remark that the dual triangular sum, associated with cubes $\ell P < \ell Q$, is estimated in a similar manner.

The family $\{(P,Q) \in \mathcal{G}_1 \times \mathcal{G}_2 \,:\, \ell P \ge \ell Q\}$ is partitioned into three subfamilies:

$$\begin{aligned}
\mathcal{P}_{\mathsf{inside}} &:= \{(P,Q) \in \mathcal{G}_1 \times \mathcal{G}_2 \;:\; Q \subset P \text{ and } 2^r \ell Q < \ell P\};\\
\mathcal{P}_{\mathsf{separated}} &:= \{(P,Q) \in \mathcal{G}_1 \times \mathcal{G}_2 \;:\; \ell Q \le \ell P \text{ and } \ell Q \le \mathrm{dist}(Q,P)\};\\
\mathcal{P}_{\mathsf{nearby}} &:= \{(P,Q) \in \mathcal{G}_1 \times \mathcal{G}_2 \;:\; 2^{-r}\ell P \le \ell Q \le \ell P \text{ and } \mathrm{dist}(Q,P) < \ell Q\}.
\end{aligned}$$

The fact that this is a partition relies on the goodness of Q. We refer to [19, Section 13] for further details. The sums over these collections of cubes are handled separately. Let us denote

$$\mathbf{B}_\star(f_1,f_2) = \sum_{(P,Q)\in\mathcal{P}_\star} \langle T\Delta_P f_1, \Delta_Q f_2\rangle\,, \qquad \star \in \{\text{inside}, \text{separated}, \text{nearby}\}.$$

The analysis of the (most difficult) inside term is performed within sections 5 and 6. It relies on a corona decomposition, and the associated stopping tree is first constructed in Section 4. The separated term is analysed in a standard manner in Section 7. Finally, throughout sections 8–10, we treat the nearby terms via surgery.

4. A stopping tree construction

A stopping tree construction is used in the analysis of the inside-term.

For $j \in \{1,2\}$, let us define a stopping tree $\mathcal{S}_j \subset \mathcal{D}_j$ and a function $\sigma_j : \mathcal{S}_j \mapsto \mathbf{R}_+$ as follows. Take the maximal good $\mathcal{D}_j$-cubes $Q \subset Q_{j,0}$ in $\mathcal{S}_j$, and define $\sigma_j(Q) := \langle |f_j| \rangle_Q$ for these maximal cubes. At inductive stage, if $S \in \mathcal{S}_j$ is a minimal cube, we consider the maximal $\mathcal{D}_j$-cubes $Q \subsetneq S$ subject to both of the conditions (1)–(2):

(1) $\langle |f_j| \rangle_Q > 4\sigma_j(S)$;
(2) Either Q or $\pi_j Q$ is a good cube.

We add these cubes Q to the stopping tree $\mathcal{S}_j$, and define $\sigma_j(Q) := \langle |f_j| \rangle_Q$ for each of them.

4.1. *Remark.* Condition (2), imposed in the construction of stopping trees, will be useful to us in many occasions. A minor side effect is that we can rely on inequality $\langle |f_j| \rangle_Q \le 4\sigma_1(\pi_{\mathcal{S}_j} Q)$ for a $\mathcal{D}_j$-dyadic cube $Q \subsetneq Q_{j,0}$ only if either Q or $\pi_j Q$ is good. But this is, in fact, all we need.

4.2. *Remark.* By construction $\mathcal{S}_j$ is a 'sparse family of cubes', i.e.,

$$\sum_{S' \in \mathrm{ch}_{\mathcal{S}_j}(S)} \mu(S') \le 4^{-1} \mu(S) , \qquad S \in \mathcal{S}_j , \quad j \in \{1,2\} . \tag{4.3}$$

In particular, family $\mathcal{S}_j$ satisfies a 'Carleson condition': $\sum_{S' \in \mathcal{S}_j : S' \subset S} \mu(S') \lesssim \mu(S)$ if $S \in \mathcal{S}_j$.

4.1. **Quasi-orthogonality.** The following is a key inequality,

$$\sum_{S \in \mathcal{S}_j} \sigma_j(S)^{p_j} \mu(S) \lesssim \|f_j\|_{p_j}^{p_j} \lesssim 1 , \qquad j \in \{1,2\} . \tag{4.4}$$

Proof of (4.4). We apply a dyadic maximal function: $M_{j,\mu} f_j(x) = \sup_{x \in Q \in \mathcal{D}_j} \langle |f_j| \rangle_Q$. For $S \in \mathcal{S}_j$, we let E_S be the set S minus all the $\mathcal{S}_j$-children of S. By inequality (4.3), $\mu(E_S) \ge \frac{3}{4}\mu(S)$, and the sets E_S are pairwise disjoint by definition. Hence,

$$\begin{aligned} \sum_{S \in \mathcal{S}_j} \sigma_j(S)^{p_j} \mu(S) &\le \tfrac{4}{3} \sum_{S \in \mathcal{S}_j} \langle |f_j| \rangle_S^{p_j} \mu(E_S) \\ &\lesssim \sum_{S \in \mathcal{S}_j} \int_{E_S} (M_{j,\mu} f_j)^{p_j} \, d\mu \le \int_{\mathbf{R}^n} (M_{j,\mu} f_j)^{p_j} \, d\mu . \end{aligned}$$

Thus, the first inequality in (4.4) follows from the fact that $M_{j,\mu}$ is bounded on L^{p_j}. The second inequality is a consequence of the martingale transform inequality (2.3). $\square$

4.2. **Martingale projections.** For $S \in \mathcal{S}_j$ and $\phi \in L^1_{\mathrm{loc}}$, we define $P_{j,S}\phi = \sum_{Q\in\mathcal{D}_j\,:\,\pi_{\mathcal{S}_j}Q=S} \Delta_Q\phi$. By orthogonality of martingale differences and inequality (2.3), for all $S \in \mathcal{S}_j$ and all sequences of constants satisfying $\sup_{Q\in\mathcal{G}_j}|\varepsilon_Q| \le 1$,

$$\Big\| \sum_{Q\in\mathcal{G}_j\,:\,\pi_{\mathcal{S}_j}Q=S} \varepsilon_Q\Delta_Q f_j \Big\|_{p_j}^{p_j} \lesssim \|P_{j,S}f_j\|_{p_j}^{p_j}\,. \tag{4.5}$$

Of fundamental importance is the following inequality, which *does not hold* for general families of orthogonal martingale projections, in the case of $1 < p_j < 2$.

$$\sum_{S\in\mathcal{S}_j} \|P_{j,S}f_j\|_{p_j}^{p_j} \lesssim 1\,. \tag{4.6}$$

Proof of (4.6). Let us write

$$\sum_{S\in\mathcal{S}_j} \|P_{j,S}f_j\|_{p_j}^{p_j} = \sum_{S\in\mathcal{S}_j} \|\mathbf{1}_{S\setminus E_S}P_{j,S}f_j\|_{p_j}^{p_j} + \sum_{S\in\mathcal{S}_j} \|\mathbf{1}_{E_S}P_{j,S}f_j\|_{p_j}^{p_j}\,, \tag{4.7}$$

where E_S denotes the set $S\setminus\bigcup_{S'\in\mathrm{ch}_{\mathcal{S}_j}(S)} S'$. We estimate the two terms separately. First,

$$|\mathbf{1}_{S\setminus E_S}P_{j,S}f_j| = \Big| \sum_{S'\in\mathrm{ch}_{\mathcal{S}_j}(S)} \mathbf{1}_{S'}\{\langle f_j\rangle_{S'} - \langle f_j\rangle_S\} \Big| \lesssim \sum_{S'\in\mathrm{ch}_{\mathcal{S}_j}(S)} \mathbf{1}_{S'}\sigma_j(S')\,.$$

Since the family $\mathrm{ch}_{\mathcal{S}_j}(S)$ is disjoint, the upper bound for the first term in RHS(4.7) follows from inequality (4.4).

By (4.4) it remains to show that $|\mathbf{1}_{E_S}P_{j,S}f_j| \lesssim \mathbf{1}_{E_S}\sigma_j(S)$ almost everywhere. We restrict ourselves to points in which $\lim_{k\to-\infty}\mathbf{E}_{j,k}f_j(x) = f_j(x)$, hence $|\mathbf{1}_{E_S}(x)P_{j,S}f_j(x)| = |\mathbf{1}_{E_S}(x)\{f_j(x) - \langle f_j\rangle_S\}|$. Observe that $|\langle f_j\rangle_S| \le \sigma_j(S)$. Now, there are three cases (1)–(3) for $x \in E_S$ as above:

(1) If there are no good $\mathcal{D}_j$-cubes inside S containing x, we have $P_{j,S}f_j(x) = 0$ by definitions.

(2) There is a minimal good $\mathcal{D}_j$-cube $Q \subset S$ containing x, in which case we let $Q_x \in \mathrm{ch}(Q)$ be the child containing x. If $R \subset Q_x$ is a $\mathcal{D}_j$-cube containing x, we easily find that $\langle f_j\rangle_R = \langle f_j\rangle_{Q_x}$. Thus, by martingale convergence,

$$|f_j(x)| = \lim_{\ell R\to 0}|\langle f_j\rangle_R| = |\langle f_j\rangle_{Q_x}| \le 4\sigma_j(\pi_{\mathcal{S}_j}Q_x) = 4\sigma_j(S)\,.$$

In the penultimate step above, we used Remark 4.1 and the fact that $\pi_j Q_x = Q$ is good. And, in the last step, we used the fact that $x \in E_S$.

(3) There are arbitrarily small good $\mathcal{D}_j$-cubes $Q \subset S$ containing x. Hence,

$$|f_j(x)| = \lim_{\ell Q\to 0}|\langle f_j\rangle_Q| \le \sup\{\langle |f_j|\rangle_Q \,:\, x \in Q \subset S\} \le 4\sigma_j(S)\,.$$

The limit and supremum above are restricted to good $\mathcal{D}_j$-cubes satisfying $x \in Q \subset S$. $\square$

4.3. **Family $\mathcal{L}_2(S)$ and its layers.** This construction is needed as we study the case of $p_j \neq 2$, and in particular it will allow us to more freely use the inequality (4.6).

For $S \in \mathcal{S}_1$ let us define $\mathcal{L}_2(S) \subset \mathcal{S}_2$ to be the family of cubes of the form $R = \pi_{\mathcal{S}_2} Q$, where $Q \in \mathcal{G}_2$ satisfies $(P, Q) \in \mathcal{P}_{\mathsf{inside}}$ for some cube $P \in \mathcal{G}_1$ with $\pi_{\mathcal{S}_1} P_Q = S$. Here P_Q stands for the child of P containing Q; it exists by goodness of Q.

Lemma 4.8 records the observation that there are at most $2(r+1)$ layers in $\mathcal{L}_2(S)$ which contain cubes R such that $R \not\subset S$. To be more precise, let $\mathcal{L}_2^k(S)$ denote the layer $k \geq 0$ cubes in $\mathcal{L}_2(S)$, i.e., the cubes R in this family for which $\pi^k_{\mathcal{L}_2(S)} R$ is a maximal cube in $\mathcal{L}_2(S)$.

4.8. **Lemma.** *Suppose that $S \in \mathcal{S}_1$ and $R \in \mathcal{L}_2^k(S)$ with $k \geq 2(r+1)$. Then $R \subset S$.*

Proof. We first claim that, if $R \in \mathcal{L}_2^k(S)$ with $k \geq 1$, then

$$2^{k-1} \ell R \leq 2^r \ell S\,. \tag{4.9}$$

The lemma is a consequence of inequality (4.9). Indeed, if $k \geq 2(r+1)$, we then have $2^{r+1} \ell R \leq \ell S$ and $S \cap R \neq \emptyset$. It remains to recall that either R or πR is a good (by construction).

Let us then prove inequality (4.9). Clearly, it suffices to verify the case of $k = 1$. Suppose that $R \subsetneq R_0$ is a cube in the first layer, and $R_0 \in \mathcal{L}_2^0(S)$ is maximal. Then, by definition, there are cubes $Q, Q' \in \mathcal{G}_2$ such that $Q \cup Q' \subset S$, $Q \subset R$ and $Q' \cap (R_0 \setminus R) \neq \emptyset$. From these facts it easily follows that $S \cap R \neq \emptyset$ and $\operatorname{dist}(S, \partial R) = 0$. Since either S or $\pi_1 S$ is a good cube, $\ell R \leq 2^r \ell S$. □

4.4. **Further inequalities.** The reader may omit this technical section for the time being. The following important inequality parallels (4.6); recall definition of $\mathcal{L}_2(S)$ in Section 4.3:—for all sequences of constants satisfying $\sup_{\mathcal{G}_2 \times \mathcal{S}_1} |\varepsilon_{Q,S}| \leq 1$,

$$\sum_{S \in \mathcal{S}_1} \sum_{S' \in \mathsf{ch}^t_{\mathcal{S}_1}(S)} \sum_{\substack{R \in \mathcal{L}_2(S) \\ R \not\subset S}} \Bigg\| \sum_{\substack{Q \in \mathcal{G}_2 : \pi_{\mathcal{S}_2} Q = R \\ \pi_{\mathcal{S}_1} Q = S'}} \varepsilon_{Q,S} \Delta_Q f_2 \Bigg\|_{p_2}^{p_2} \lesssim 1\,, \qquad \text{if } t \geq 0\,. \tag{4.10}$$

Proof of inequality (4.10). Let us fix $t \geq 0$. First, by martingale transform inequality (2.3) and orthogonality of martingale differences, we can assume that $\varepsilon_{Q,S} = 1$ for all Q and S. By Lemma 4.12, we obtain an upper bound

$$\sum_{R \in \mathcal{S}_2} \sigma_2(R)^{p_2} \sum_{S \in \mathcal{S}_1; R \not\subset S} \mu(S \cap R) + \sum_{R \in \mathcal{S}_2} \sum_{R' \in \mathsf{ch}_{\mathcal{S}_2}(R)} \sum_{S \in \mathcal{S}_1} \sum_{\substack{S' \in \mathsf{ch}^t_{\mathcal{S}_1}(S) \\ \pi_{\mathcal{S}_1}(\pi_2 R') = S'}} \sigma_2(R')^{p_2} \mu(R')\,. \tag{4.11}$$

By inequality (4.4), the second term is bounded by $\lesssim 1$; Indeed, for a fixed R' there is at most one pair of cubes S, S' such that $\pi_{\mathcal{S}_1}(\pi_2 R') = S'$.

Concerning the first term in (4.11), we observe that $\sum_{S\in\mathcal{S}_1;R\not\subset S}\mu(S\cap R)\lesssim\mu(R)$ and then apply inequality (4.4). Mentioned observation is reached by splitting the series in two parts, depending if $S\subset R$ or not; The series with $S\subset R$ is estimated by the Carleson condition, Remark 4.2. The second series, in which $S\not\subset R$, is estimated by using the fact that $2^{-r}\ell S\le\ell R\le 2^r\ell S$ if $S\cap R\neq\emptyset$, $S\not\subset R$, and $R\not\subset S$. Indeed, there are at most $c(n,r)$ such cubes $S\in\mathcal{S}_1$ for a fixed $R\in\mathcal{S}_2$. □

4.12. **Lemma.** *Let us fix $S\in\mathcal{S}_1$, $S'\in\mathrm{ch}^t_{\mathcal{S}_1}(S)$, and $R\in\mathcal{L}_2(S)$ such that $R\not\subset S$. Then*

$$\Bigg\|\sum_{\substack{Q\in\mathcal{G}_2:\pi_{\mathcal{S}_2}Q=R\\ \pi_{\mathcal{S}_1}Q=S'}}\Delta_Q f_2\Bigg\|_{p_2}^{p_2}\lesssim\mu(S'\cap R)\sigma_2(R)^{p_2}+\sum_{R'\in\mathrm{ch}_{\mathcal{S}_2}(R)}\mathbf{1}_{\pi_{\mathcal{S}_1}(\pi_2R')=S'}\mu(R')\sigma_2(R')^{p_2}\,. \tag{4.13}$$

Proof. Let E_R be the set R take away all the $\mathcal{S}_2$-children of R. Then, LHS(4.13) is bounded by

$$\mathbf{A}+\mathbf{B}=\Bigg\|\mathbf{1}_{R\setminus E_R}\sum_{\substack{Q\in\mathcal{G}_2:\pi_{\mathcal{S}_2}Q=R\\ \pi_{\mathcal{S}_1}Q=S'}}\Delta_Q f_2\Bigg\|_{p_2}^{p_2}+\Bigg\|\mathbf{1}_{E_R\cap S'}\sum_{\substack{Q\in\mathcal{G}_2:\pi_{\mathcal{S}_2}Q=R\\ \pi_{\mathcal{S}_1}Q=S'}}\Delta_Q f_2\Bigg\|_{p_2}^{p_2}.$$

In order to estimate $\mathbf{A}$, let us consider a child $R'\in\mathrm{ch}_{\mathcal{S}_2}(R)$ for which there are cubes $Q^+=Q^+(R')$ and $Q^-=Q^-(R')$:—these are the maximal and minimal cubes, respectively, subject to conditions $Q\in\mathcal{G}_2$, $\pi_{\mathcal{S}_2}Q=R$, $\pi_{\mathcal{S}_1}Q=S'$, and $R'\subsetneq Q$. Then

$$\begin{aligned}\Bigg|\mathbf{1}_{R'}\sum_{\substack{Q\in\mathcal{G}_2:\pi_{\mathcal{S}_2}Q=R\\ \pi_{\mathcal{S}_1}Q=S'}}\Delta_Q f_2\Bigg|&=\mathbf{1}_{R'\cap S'}\cdot|\{\langle f_2\rangle_{Q^-_{R'}}-\langle f_2\rangle_{Q^+}\}|\\ &\lesssim\begin{cases}\mathbf{1}_{\pi_{\mathcal{S}_1}(\pi_2R')=S'}\mathbf{1}_{R'}\sigma_2(R')\,, & \text{if } Q^-_{R'}=R';\\ \mathbf{1}_{S'\cap R'}\sigma_2(R)\,, & \text{otherwise}\,.\end{cases}\end{aligned}$$

We used the facts that $\Delta_Q f_2=0$ if $Q\subset Q_{2,0}$ is bad, and that $\pi_{\mathcal{S}_2}Q^-_{R'}=R$ if $Q^-_{R'}\neq R'$. Writing $R\setminus E_R=\sum_{R'\in\mathrm{ch}_{\mathcal{S}_2}(R)}\mathbf{1}_{R'}$ and using disjointness of these children yields

$$\mathbf{A}\lesssim\mu(S'\cap R)\sigma_2(R)^{p_2}+\sum_{R'\in\mathrm{ch}_{\mathcal{S}2}(R)}\mathbf{1}_{\pi_{\mathcal{S}_1}(\pi_2R')=S'}\mu(R')\sigma_2(R')^{p_2}\,.$$

We turn to term $\mathbf{B}$; we will implicitly use the fact that $\Delta_Q f_2=0$ if $Q\subset Q_{2,0}$ is a bad $\mathcal{D}_2$-cube. Let us fix a point $x\in E_R\cap S'$ such that $\lim_{k\to-\infty}\mathbf{E}_{2,k}f_2(x)=f_2(x)$, and there is a maximal cube $Q^+\in\mathcal{G}_2$, subject to conditions $x\in Q\in\mathcal{G}_2$, $\pi_{\mathcal{S}_2}Q=R$, $\pi_{\mathcal{S}_1}Q=S'$. Then,

$$\Bigg|\sum_{\substack{Q\in\mathcal{G}_2:\pi_{\mathcal{S}_2}Q=R\\ \pi_{\mathcal{S}_1}Q=S'}}\Delta_Q f_2(x)\Bigg|=\Bigg|\sum_{\substack{Q\in\mathcal{G}_2:\pi_{\mathcal{S}_1}Q=S'\\ Q\subset Q^+}}\Delta_Q f_2(x)\Bigg|. \tag{4.14}$$

We aim to verify that RHS(4.14) $\lesssim \sigma_2(R)$. This allows us to conclude that $\mathbf{B} \lesssim \mu(S' \cap R)\sigma_2(R)^{p_2}$. There are two cases. First, $\pi_{\mathcal{S}_1} Q = S'$ for all cubes $x \in Q \in \mathcal{G}_2$ with $Q \subset Q^+$; In this case, we proceed as in the proof of (4.6) in order to see that RHS(4.14) $= |f_2(x) - \langle f_2 \rangle_{Q^+}| \lesssim \sigma_2(R)$. Second, there is a minimal cube Q^- subject to conditions $\pi_{\mathcal{S}_1} Q = S'$, $x \in Q \in \mathcal{G}_2$, $Q \subset Q^+$. In this case, we find that RHS(4.14) $= |\langle f_2 \rangle_{Q_x^-} - \langle f_2 \rangle_{Q^+}| \lesssim \sigma_2(R)$, where Q_x^- denotes the child of Q^-, containing x. In the last step, we used the fact that $x \in E_R$ so that $\pi_{\mathcal{S}_2}(Q_x^-) = R$. □

5. The Inside-Paraproduct Term

First we decompose the inside term $\mathbf{B}_{\text{inside}}(f_1, f_2)$, associated with the indexing cubes $\mathcal{P}_{\text{inside}}$. There will be three terms labelled as: 'paraproduct', 'stopping', and 'error'. The 'paraproduct' term is treated in this section. The other ones are treated in Section 6.

The conditions for $(P, Q) \in \mathcal{P}_{\text{inside}}$ are: $P \in \mathcal{G}_1$, $Q \in \mathcal{G}_2$, $Q \subset P$, and $2^r \ell Q < \ell P$. These are abbreviated to $Q \Subset P$. The child of P containing Q is denoted by P_Q; it exists by goodness of Q. For $(P, Q) \in \mathcal{P}_{\text{inside}}$ we write

$$\Delta_P f_1 = \langle \Delta_P f_1 \rangle_{P_Q} \mathbf{1}_{\pi_{\mathcal{S}_1} P_Q} - \langle \Delta_P f_1 \rangle_{P_Q} \mathbf{1}_{\pi_{\mathcal{S}_1} P_Q \setminus P_Q} + \Delta_P f_1 \cdot \mathbf{1}_{P \setminus P_Q} .$$

This equation is valid pointwise μ-almost everywhere (everywhere if $\mu(P_Q) \neq 0$), and it yields the following expansion, respectively,

$$\begin{aligned} \mathbf{B}_{\text{inside}}(f_1, f_2) &= \sum_{(P,Q) \in \mathcal{P}_{\text{inside}}} \langle T \Delta_P f_1, \Delta_Q f_2 \rangle \\ &= \mathbf{B}^{\text{para}}(f_1, f_2) - \mathbf{B}^{\text{stop}}(f_1, f_2) + \mathbf{B}^{\text{error}}(f_1, f_2) , \end{aligned}$$

Hence, e.g., $\mathbf{B}^{\text{para}}(f_1, f_2) = \sum_{(P,Q) \in \mathcal{P}_{\text{inside}}} \langle \Delta_P f_1 \rangle_{P_Q} \langle T \mathbf{1}_{\pi_{\mathcal{S}_1} P_Q}, \Delta_Q f_2 \rangle$. The main result in this section is the following estimate for the paraproduct term.

5.1. **Proposition.** *We have inequality* $\left| \mathbf{B}^{\text{para}}(f_1, f_2) \right| \lesssim 1 + \mathbf{T}_{\text{loc}}$.

The remainder of this section is dedicated to the proof of this proposition, and the main focus will be on auxiliary inequalities (5.6) and (5.11). Let us first examine how these inequalities are used to prove Proposition 5.1. First, for $S \in \mathcal{S}_1$, recall definition of $\mathcal{L}_2(S)$ given in Section 4.3. We define

$$\begin{aligned} &\mathbf{B}^{\text{para}}_{S,\not\subset}(f_1, f_2) + \mathbf{B}^{\text{para}}_{S,\subset}(f_1, f_2) \\ &= \Bigg\{ \sum_{\substack{R \in \mathcal{L}_2(S) \\ R \not\subset S}} + \sum_{\substack{R \in \mathcal{L}_2(S) \\ R \subset S}} \Bigg\} \sum_{Q \in \mathcal{G}_2 : \pi_{\mathcal{S}_2} Q = R} \sum_{\substack{P \in \mathcal{G}_1 : \pi_{\mathcal{S}_1} P_Q = S \\ Q \Subset P}} \langle \Delta_P f_1 \rangle_{P_Q} \langle T \mathbf{1}_S, \Delta_Q f_2 \rangle . \end{aligned}$$

Then, by the auxiliary inequalities mentioned above,

$$|\mathbf{B}^{\mathrm{para}}(f_1,f_2)| \le \Big|\sum_{S\in\mathcal{S}_1}\mathbf{B}^{\mathrm{para}}_{S,\not\subset}(f_1,f_2)\Big| + \Big|\sum_{S\in\mathcal{S}_1}\mathbf{B}^{\mathrm{para}}_{S,\subset}(f_1,f_2)\Big| \lesssim 1+\mathbf{T}_{\mathrm{loc}}\,.$$

This concludes the proof of Proposition 5.1, assuming the auxiliary inequalities.

5.1. **A telescoping identity.** For fixed $Q\in\mathcal{G}_2$ and $S\in\mathcal{S}_1$, let us define a constant $\varepsilon_{Q,S}$ by

$$\varepsilon_{Q,S}\sigma_1(S) = \sum_{\substack{P\in\mathcal{G}_1:\pi_{\mathcal{S}_1}P_Q=S\\ Q\Subset P}} \langle \Delta_P f_1\rangle_{P_Q}\,. \tag{5.2}$$

It is important to use the condition $\pi_{\mathcal{S}_1}P_Q=S$ instead of $\pi_{\mathcal{S}_1}P=S$. Otherwise, the following important lemma might fail, as the measure μ need not be doubling.

5.3. **Lemma.** *For $Q\in\mathcal{G}_2$ and $S\in\mathcal{S}_1$, we have $|\varepsilon_{Q,S}|\lesssim 1$.*

Proof. Recall our convention that $\langle\Delta_P f_1\rangle_{P_Q}=0$ if $\mu(P_Q)=0$. Consider the minimal and maximal dyadic cubes: P^- and P^+, subject to conditions $P\in\mathcal{G}_1$, $\pi_{\mathcal{S}_1}P_Q=S$, $Q\Subset P$, and $\mu(P_Q)\neq 0$. If such cubes do not exist, we are done. Otherwise, we claim that

$$\varepsilon_{Q,S}\sigma_1(S) = \langle f_1\rangle_{P^-_Q} - \langle f_1\rangle_{P^+}\,. \tag{5.4}$$

By using equation (5.4) and the construction of the stopping tree, we find that $|\varepsilon_{Q,S}|\le 8$.

It remains to prove equation (5.4). Suppose that $P\in\mathcal{G}_1$ is such that $\pi_{\mathcal{S}_1}P_Q=S$, $Q\Subset P$, and $\mu(P_Q)\neq 0$. Then $P^-\subset P\subset P^+$. By this observation,

$$\sum_{\substack{P\in\mathcal{G}_1:\pi_{\mathcal{S}_1}P_Q=S\\ Q\Subset P:\mu(P_Q)\neq 0}} \Delta_P f_1\cdot \mathbf{1}_{P^-_Q} = \sum_{\substack{P\in\mathcal{G}_1\\ P^-\subset P\subset P^+}} \mathbf{1}_{\pi_{\mathcal{S}_1}P_Q=S}\mathbf{1}_{Q\Subset P}\mathbf{1}_{\mu(P_Q)\neq 0}\cdot\Delta_P f_1\cdot\mathbf{1}_{P^-_Q}\,. \tag{5.5}$$

Observe that $\mathbf{1}_{\pi_{\mathcal{S}_1}P_Q=S}\mathbf{1}_{Q\Subset P}\mathbf{1}_{\mu(P_Q)\neq 0}=1$ inside the summation. Also, $\Delta_P f_1=0$ if P is a bad cube with $P^-\subset P\subset P^+$. Thus, by adding the zero contribution from the bad cubes in a formal manner, we obtain a telescoping identity: $\mathrm{LHS}(5.5) = \{\langle f_1\rangle_{P^-_Q}-\langle f_1\rangle_{P^+}\}\mathbf{1}_{P^-_Q}$. The equation (5.4) follows from this: first, we restrict ourselves to cubes P with $\mu(P_Q)\neq 0$ in the series defining $\varepsilon_{Q,S}$. Then, we replace the P_Q averages by P^-_Q averages inside the summation; observe that $P^-_Q\subset P_Q$ and $\mu(P^-_Q)\neq 0$. Finally, we exchange the order of summation and the brackets, and apply the obtained telescoping identity. □

5.2. **Summation involving cubes** $R \not\subset S$**.** Our aim in this section is to prove an inequality,

$$\Big|\sum_{S\in\mathcal{S}_1} \mathbf{B}^{\mathsf{para}}_{S,\not\subset}(f_1,f_2)\Big| \lesssim 1+\mathbf{T}_{\mathsf{loc}}\,. \tag{5.6}$$

Let us express the series defining $\mathbf{B}^{\mathsf{para}}_{S,\not\subset}(f_1,f_2)$ in a convenient manner. For this purpose, observe that $Q\subset P_Q\subset S$ for any cube Q in the series defining $\mathbf{B}^{\mathsf{para}}_{S,\not\subset}(f_1,f_2)$. In particular, $\pi_{\mathcal{S}_1}Q\subset S$. Thus, by organising the Q-summation in terms of their $\mathcal{S}_1$-parents and defining $\varepsilon_{Q,S}$ as the solution to equation (5.2), we find that

$$\mathbf{B}^{\mathsf{para}}_{S,\not\subset}(f_1,f_2) = \sigma_1(S)\sum_{t\ge 0}\sum_{\substack{R\in\mathcal{L}_2(S)\\ R\not\subset S}}\sum_{S'\in \mathsf{ch}^t_{\mathcal{S}_1}(S)}\sum_{\substack{Q\in\mathcal{G}_2:\pi_{\mathcal{S}_2}Q=R\\ \pi_{\mathcal{S}_1}Q=S'}}\langle \mathbf{1}_R\mathbf{1}_{S'}\{T\mathbf{1}_S-\tau_{t,S'}\},\varepsilon_{Q,S}\Delta_Q f_2\rangle\,. \tag{5.7}$$

By using the fact that $\Delta_Q f_2$ has mean zero, we have also subtracted off the constants

$$\tau_{t,S'} = \begin{cases} 0\,, & \text{if } t\in\{0,\dots,2r+1\};\\ T\mathbf{1}_{S\setminus \pi^{\lfloor t/2\rfloor}_{\mathcal{S}_1}S'}(x_{S'})\,, & \text{otherwise}\,.\end{cases}$$

For convenience, let us denote

$$\mathbf{A}_{t,S} = \Big\{\sum_{\substack{R\in\mathcal{L}_2(S)\\ R\not\subset S}}\sum_{S'\in \mathsf{ch}^t_{\mathcal{S}_1}(S)} \|\mathbf{1}_R\mathbf{1}_{S'}\{T\mathbf{1}_S-\tau_{t,S'}\}\|_{p_1}^{p_1}\Big\}^{1/p_1}\,,$$

and

$$\mathbf{B}_{t,S} = \Big\{\sum_{\substack{R\in\mathcal{L}_2(S)\\ R\not\subset S}}\sum_{S'\in \mathsf{ch}^t_{\mathcal{S}_1}(S)} \Big\|\sum_{\substack{Q\in\mathcal{G}_2:\pi_{\mathcal{S}_2}Q=R\\ \pi_{\mathcal{S}_1}Q=S'}} \varepsilon_{Q,S}\Delta_Q f_2\Big\|_{p_2}^{p_2}\Big\}^{1/p_2}\,.$$

The useful inequality $\sup_{\mathcal{G}_2\times\mathcal{S}_1}|\varepsilon_{Q,S}|\lesssim 1$ is a consequence of Lemma 5.3. By equation (5.7) and Hölder's inequality, combined with Lemma 5.8,

$$\begin{aligned}\Big|\sum_{S\in\mathcal{S}_1}\mathbf{B}^{\mathsf{para}}_{S,\not\subset}(f_1,f_2)\Big| &\lesssim \sum_{t\ge0}\sum_{S\in\mathcal{S}_1}\sigma_1(S)\mathbf{A}_{t,S}\mathbf{B}_{t,S}\\ &\lesssim (1+\mathbf{T}_{\mathsf{loc}})\sum_{t\ge 0}2^{-t/p_1}\Big\{\sum_{S\in\mathcal{S}_1}\sigma_1(S)^{p_1}\mu(S)\Big\}^{1/p_1}\Big\{\sum_{S\in\mathcal{S}_1}\mathbf{B}_{t,S}^{p_2}\Big\}^{1/p_2}\,.\end{aligned}$$

Inequality (5.6) is obtained by applying inequalities (4.4) and (4.10), and summing the geometric series afterwards.

5.8. **Lemma.** *For every $t\ge 0$ and $S\in\mathcal{S}_1$, we have $\mathbf{A}_{t,S}\lesssim(1+\mathbf{T}_{\mathsf{loc}})2^{-t/p_1}\mu(S)^{1/p_1}$.*

Proof. By Lemma 4.8 and the fact that layers $\mathcal{L}_2^k(S)$, $k \geq 0$, are comprised of disjoint cubes, we can bound $\mathbf{A}_{t,S}^{p_1}$ by

$$\sum_{k=0}^{2r+1} \sum_{S' \in \mathrm{ch}_{\mathcal{S}_1}^t(S)} \sum_{\substack{R \in \mathcal{L}_2^k(S) \\ R \not\subset S}} \|\mathbf{1}_R \mathbf{1}_{S'} \{T\mathbf{1}_S - \tau_{t,S'}\}\|_{p_1}^{p_1} \lesssim \sum_{S' \in \mathrm{ch}_{\mathcal{S}_1}^t(S)} \|\mathbf{1}_{S'} \{T\mathbf{1}_S - \tau_{t,S'}\}\|_{p_1}^{p_1} . \tag{5.9}$$

Let us first focus on the case of $t \in \{0, \dots, 2r+1\}$. By inequality (5.9) and the facts that cubes in $\mathrm{ch}_{\mathcal{S}_1}^t(S)$ are disjoint and they are contained in S,

$$\mathbf{A}_{t,S}^{p_1} \lesssim \|\mathbf{1}_S T \mathbf{1}_S\|_{p_1}^{p_1} \leq \mathbf{T}_{\mathrm{loc}}^{p_1} \mu(S) \lesssim (1 + \mathbf{T}_{\mathrm{loc}})^{p_1} 2^{-t} \mu(S) .$$

Let us then focus on the case of $t \geq 2r+2$; we begin by writing

$$\mathrm{RHS}(5.9) = \sum_{S'' \in \mathrm{ch}_{\mathcal{S}_1}^{\lceil t/2 \rceil}(S)} \sum_{\substack{S' \in \mathrm{ch}_{\mathcal{S}_1}^t(S) \\ \pi_{\mathcal{S}_1}^{\lfloor t/2 \rfloor} S' = S''}} \|\mathbf{1}_{S'} \{T\mathbf{1}_S - T\mathbf{1}_{S \setminus S''}(x_{S'})\}\|_{p_1}^{p_1} .$$

To conclude the proof of lemma, it suffices to first verify that for all $S'' \in \mathrm{ch}_{\mathcal{S}_1}^{\lceil t/2 \rceil}(S)$,

$$\sum_{\substack{S' \in \mathrm{ch}_{\mathcal{S}_1}^t(S) \\ \pi_{\mathcal{S}_1}^{\lfloor t/2 \rfloor} S' = S''}} \|\mathbf{1}_{S'} \{T\mathbf{1}_S - T\mathbf{1}_{S \setminus S''}(x_{S'})\}\|_{p_1}^{p_1} \lesssim (1 + \mathbf{T}_{\mathrm{loc}})^{p_1} \mu(S'') , \tag{5.10}$$

and then inductively apply the sparseness property of $\mathcal{S}_1$, we refer to Remark 4.2.

In order to prove the remaining inequality (5.10), we estimate LHS(5.10) by $2^{p_1 - 1}(\alpha + \beta)$,

$$\alpha + \beta = \sum_{\substack{S' \in \mathrm{ch}_{\mathcal{S}_1}^t(S) \\ \pi_{\mathcal{S}_1}^{\lfloor t/2 \rfloor} S' = S''}} \|\mathbf{1}_{S'} T \mathbf{1}_{S''}\|_{p_1}^{p_1} + \sum_{\substack{S' \in \mathrm{ch}_{\mathcal{S}_1}^t(S) \\ \pi_{\mathcal{S}_1}^{\lfloor t/2 \rfloor} S' = S''}} \|\mathbf{1}_{S'} \{T\mathbf{1}_{S \setminus S''} - T\mathbf{1}_{S \setminus S''}(x_{S'})\}\|_{p_1}^{p_1} .$$

Observe that the cubes S' are contained in S'', and they are disjoint. The Local Testing Condition implies the inequality $\alpha \leq \mathbf{T}_{\mathrm{loc}}^{p_1} \mu(S'')$. In order to analyse term β, we fix $S' \in \mathrm{ch}_{\mathcal{S}_1}^t(S)$ such that $\pi_{\mathcal{S}_1}^{\lfloor t/2 \rfloor} S' = S''$. Since $\lfloor t/2 \rfloor \geq r+1$, we have $2^r \ell S' < \ell S''$. By construction of the stopping cubes, either S' or $\pi_1 S'$ is good. In both of these cases, by goodness[1], we have $\ell S' \leq \mathrm{dist}(S', \partial S'')$. Hence, by the off-diagonal estimate (2.7), we have $|T\mathbf{1}_{S \setminus S''}(x) - T\mathbf{1}_{S \setminus S''}(x_{S'})| \lesssim 1$ if $x \in S'$. This inequality allows us to conclude that $\beta \lesssim \mu(S'')$. □

[1] This application is the principal motivation for our definition of goodness; recall that good cubes are neither 1-bad nor 2-bad. The same application arises also later, Lemma 5.12.

5.3. **Summation involving cubes** $R \subset S$**.** Here we show the inequality,

$$\left| \sum_{S \in \mathcal{S}_1} \mathbf{B}^{\mathsf{para}}_{S,\subset}(f_1, f_2) \right| \lesssim 1 + \mathbf{T}_{\mathsf{loc}} \,. \tag{5.11}$$

Let us fix $S \in \mathcal{S}_1$, and express the series defining $\mathbf{B}^{\mathsf{para}}_{S,\subset}(f_1, f_2)$ in a convenient manner. For a cube $S' \in \mathcal{S}_1$, we denote by $\mathcal{R}(S')$ the family of maximal cubes in $\{R' \in \mathcal{S}_2 : \pi_{\mathcal{S}_1} R' = S'\}$; this can be an empty family. By defining constants $\varepsilon_{Q,S}$ as solutions to (5.2), we can write $\mathbf{B}^{\mathsf{para}}_{S,\subset}(f_1, f_2)$ as

$$\sigma_1(S) \sum_{t,k \geq 0} \sum_{S' \in \mathsf{ch}^t_{\mathcal{S}_1}(S)} \sum_{R \in \mathcal{R}(S')} \sum_{\substack{R' \in \mathsf{ch}^k_{\mathcal{S}_2}(R) \\ \pi_{\mathcal{S}_1} R' = S'}} \sum_{Q \in \mathcal{G}_2 : \pi_{\mathcal{S}_2} Q = R'} \langle \mathbf{1}_{R'} \{ T\mathbf{1}_S - \tau_{t,k,S',R'} \}, \varepsilon_{Q,S} \Delta_Q f_2 \rangle \,,$$

where we have denoted

$$\tau_{t,k,S',R'} = \begin{cases} 0\,, & \text{if } t, k \in \{0, \ldots, 2r+1\}; \\ T\mathbf{1}_{S \setminus \pi^{\lfloor k/2 \rfloor}_{\mathcal{S}_2} R'}(x_{R'})\,, & \text{if } k \geq 2(r+1)\,; \\ T\mathbf{1}_{S \setminus \pi^{\lfloor t/2 \rfloor}_{\mathcal{S}_1} S'}(x_{S'})\,, & \text{otherwise}\,. \end{cases}$$

It will be convenient to denote for all $t \geq 0$,

$$\mathbf{A}_{t,S} = \left\{ \sum_{k \geq 0} \sum_{S' \in \mathsf{ch}^t_{\mathcal{S}_1}(S)} \sum_{R \in \mathcal{R}(S')} \sum_{\substack{R' \in \mathsf{ch}^k_{\mathcal{S}_2}(R) \\ \pi_{\mathcal{S}_1} R' = S'}} \| \mathbf{1}_{R'} \{ T\mathbf{1}_S - \tau_{t,k,S',R'} \} \|_{p_1}^{p_1} \right\}^{1/p_1} .$$

The useful inequality $\sup_{\mathcal{G}_2 \times \mathcal{S}_1} |\varepsilon_{Q,S}| \lesssim 1$ is a consequence of Lemma 5.3. Hence, by Lemma 5.12 and Hölder's inequality, combined with inequality (4.5),

$$\begin{aligned} |\mathbf{B}^{\mathsf{para}}_{S,\subset}(f_1, f_2)| &\lesssim \sum_{t \geq 0} \sigma_1(S) \mathbf{A}_{t,S} \left\{ \sum_{S' \in \mathsf{ch}^t_{\mathcal{S}_1}(S)} \sum_{\substack{R' \in \mathcal{S}_2 \\ \pi_{\mathcal{S}_1} R' = S'}} \left\| \sum_{Q \in \mathcal{G}_2 : \pi_{\mathcal{S}_2} Q = R'} \varepsilon_{Q,S} \Delta_Q f_2 \right\|_{p_2}^{p_2} \right\}^{1/p_2} \\ &\lesssim (1 + \mathbf{T}_{\mathsf{loc}}) \sum_{t \geq 0} 2^{-t/p_1} \sigma_1(S) \mu(S)^{1/p_1} \left\{ \sum_{S' \in \mathsf{ch}^t_{\mathcal{S}_1}(S)} \sum_{\substack{R' \in \mathcal{S}_2 \\ \pi_{\mathcal{S}_1} R' = S'}} \| P_{2,R'} f_2 \|_{p_2}^{p_2} \right\}^{1/p_2} . \end{aligned}$$

The very last upper bound is summable in $S \in \mathcal{S}_1$. Indeed, after changing the order of S and t summations, an application of inequalities (4.4) and (4.6) leaves us a geometric series in t. The proof of inequality (5.11) is complete.

5.12. **Lemma.** *For each* $S \in \mathcal{S}_1$ *and* $t \geq 0$*, we have* $\mathbf{A}_{t,S} \lesssim (1 + \mathbf{T}_{\mathsf{loc}}) 2^{-t/p_1} \mu(S)^{1/p_1}$.

Proof. Let us make a case study, and first assume that $t \in \{0, \dots, 2r+1\}$. We split $\mathbf{A}^{p_1}_{t,S}$ in two subseries, subject to $k \in \{0, \dots, 2r+1\}$ and $k \geq 2(r+1)$. For a fixed $k \in \{0, \dots, 2r+1\}$, we rely on disjointness properties of layers and maximal cubes in order to see that

$$\sum_{S' \in \mathrm{ch}^t_{\mathcal{S}_1}(S)} \sum_{R \in \mathcal{R}(S')} \sum_{\substack{R' \in \mathrm{ch}^k_{\mathcal{S}_2}(R) \\ \pi_{\mathcal{S}_1} R' = S'}} \|\mathbf{1}_{R'}\{T\mathbf{1}_S - \tau_{t,k,S',R'}\}\|^{p_1}_{p_1} \leq \|\mathbf{1}_S T\mathbf{1}_S\|^{p_1}_{p_1} \leq \mathbf{T}^{p_1}_{\mathrm{loc}} \mu(S)\,.$$

Applying these inequalities with finite number of indices $k \in \{0, \dots, 2r+1\}$ shows the required inequality for the first subseries. The second subseries is bounded by

$$\begin{aligned} &\sum_{k \geq 2r+2} \sum_{S' \in \mathrm{ch}^t_{\mathcal{S}_1}(S)} \sum_{R \in \mathcal{R}(S')} \sum_{R'' \in \mathrm{ch}^{\lceil k/2 \rceil}_{\mathcal{S}_2}(R)} \sum_{\substack{R' \in \mathrm{ch}^k_{\mathcal{S}_2}(R) \\ \pi^{\lfloor k/2 \rfloor}_{\mathcal{S}_2} R' = R''}} \|\mathbf{1}_{R'}\{T\mathbf{1}_S - T\mathbf{1}_{S \setminus R''}(x_{R'})\}\|^{p_1}_{p_1} \\ &\qquad \lesssim (1+\mathbf{T}_{\mathrm{loc}})^{p_1} \sum_{k \geq 2r+2} 2^{-k} \sum_{S' \in \mathrm{ch}^t_{\mathcal{S}_1}(S)} \mu(S') \lesssim (1+\mathbf{T}_{\mathrm{loc}})^{p_1} \mu(S)\,. \end{aligned} \tag{5.13}$$

In the first step above, we applied a simple modification of inequality (5.10) and sparsness property of $\mathcal{S}_2$, we refer to Remark 4.2.

Let us then focus on the case of $t \geq 2(r+1)$. Again, we split the series $\mathbf{A}^{p_1}_{t,S}$ in two subseries as before. For the first subseries, associated with indices $k \in \{0, \dots, 2r+1\}$, we use inequality

$$\begin{aligned} &\sum_{S' \in \mathrm{ch}^t_{\mathcal{S}_1}(S)} \sum_{R \in \mathcal{R}(S')} \sum_{\substack{R' \in \mathrm{ch}^k_{\mathcal{S}_2}(R) \\ \pi_{\mathcal{S}_1} R' = S'}} \|\mathbf{1}_{R'}\{T\mathbf{1}_S - \tau_{t,k,S',R'}\}\|^{p_1}_{p_1} \\ &\qquad \leq \sum_{S'' \in \mathrm{ch}^{\lceil t/2 \rceil}_{\mathcal{S}_1}(S)} \sum_{\substack{S' \in \mathrm{ch}^t_{\mathcal{S}_1}(S) \\ \pi^{\lfloor t/2 \rfloor}_{\mathcal{S}_1} S' = S''}} \|\mathbf{1}_{S'}\{T\mathbf{1}_S - T\mathbf{1}_{S \setminus S''}(x_{S'})\}\|^{p_1}_{p_1}\,, \end{aligned}$$

and then proceed as in the proof of Lemma 5.8. Finally, the second subseries is bounded by LHS(5.13) which, in turn, is controlled by $\lesssim (1+\mathbf{T}_{\mathrm{loc}})^{p_1} 2^{-t} \mu(S)$, Remark 4.2. □

6. The Inside-Stopping/Error Term

In the present section, we concentrate on the two terms, labelled as 'stopping' and 'error', that were introduced in the beginning of Section 5. We aim to prove the following proposition.

6.1. **Proposition.** *We have* $\left|\mathbf{B}^{\mathrm{stop}}(f_1, f_2)\right| + \left|\mathbf{B}^{\mathrm{error}}(f_1, f_2)\right| \lesssim 1$.

6.1. **The stopping term.** The stopping term $\mathbf{B}^{\mathsf{stop}}(f_1,f_2)$ is written as $\sum_{t=r+1}^{\infty}\mathbf{B}_t^{\mathsf{stop}}(f_1,f_2)$,

$$|\mathbf{B}_t^{\mathsf{stop}}(f_1,f_2)| = \Bigg|\sum_{P\in\mathcal{G}_1}\sum_{\substack{Q\in\mathcal{G}_2\\ 2^t\ell Q=\ell P}} \mathbf{1}_{Q\subset P}\langle\Delta_P f_1\rangle_{P_Q}\langle T\mathbf{1}_{\pi_{\mathcal{S}_1}P_Q\setminus P_Q},\Delta_Q f_2\rangle\Bigg|$$

$$\lesssim 2^{-t\eta(1-\gamma)}\int_{\mathbf{R}^n}\sum_{P\in\mathcal{G}_1}\sum_{\substack{Q\in\mathcal{G}_2\\ 2^t\ell Q=\ell P}} \mathbf{1}_Q(x)|\Delta_P f_1(x)|\cdot\mathbf{1}_{Q\subset P}|\Delta_Q f_2(x)|\,d\mu(x)\,.$$

In the last step, we used the off-diagonal estimate (2.7) and the fact that $\Delta_Q f_2$ has mean zero. Applying Cauchy–Schwarz and Hölder's inequality, and then observing inequalities,

$$\sum_{\substack{P\in\mathcal{G}_1\\ 2^t\ell Q=\ell P}}\mathbf{1}_{Q\subset P}\le 1\quad(Q\in\mathcal{G}_2)\,,\qquad \sum_{\substack{Q\in\mathcal{G}_2\\ 2^t\ell Q=\ell P}}\mathbf{1}_Q\le\mathbf{1}_{\mathbf{R}^n}\quad(P\in\mathcal{G}_1)\,,$$

we obtain, for a fixed $t\ge r+1$,

$$|\mathbf{B}_t^{\mathsf{stop}}(f_1,f_2)|\lesssim 2^{-t\eta(1-\gamma)}\Bigg\|\Bigg(\sum_{P\in\mathcal{G}_1}|\Delta_P f_1|^2\Bigg)^{1/2}\Bigg\|_{p_1}\Bigg\|\Bigg(\sum_{Q\in\mathcal{G}_2}|\Delta_Q f_2|^2\Bigg)^{1/2}\Bigg\|_{p_2}\lesssim 2^{-t\eta(1-\gamma)}\,.$$

In the penultimate step, we used inequality (2.4). The last bound is summable in t, and this concludes analysis of the stopping term.

6.2. **The error term.** We write $\mathbf{B}^{\mathsf{error}}(f_1,f_2)=\sum_{t=r+1}^{\infty}\mathbf{B}_t^{\mathsf{error}}(f_1,f_2)$,

$$\mathbf{B}_t^{\mathsf{error}}(f_1,f_2)=\sum_{j=1}^{2^n}\sum_{Q\in\mathcal{G}_2}\sum_{\substack{P\in\mathcal{G}_1\\ 2^t\ell Q=\ell P}}\mathbf{1}_{Q\subset P}\mathbf{1}_{P_j\neq P_Q}\langle\Delta_P f_1\rangle_{P_j}\langle T\mathbf{1}_{P_j},\Delta_Q f_2\rangle\,.$$

Let us denote $T_{P_j,Q}=\mathbf{1}_{P_j\neq P_Q}\{T\mathbf{1}_{P_j}-T\mathbf{1}_{P_j}(x_Q)\}$. By the fact that $\Delta_Q f_2$ has mean zero, we can bound $|\mathbf{B}_t^{\mathsf{error}}(f_1,f_2)|$ with $t\ge r+1$ by

$$\sum_{j=1}^{2^n}\Bigg|\int_{\mathbf{R}^n}\sum_{Q\in\mathcal{G}_2}\Delta_Q f_2(x)\cdot\mathbf{1}_Q(x)\sum_{\substack{P\in\mathcal{G}_1\\ 2^t\ell Q=\ell P}}\mathbf{1}_{Q\subset P}\mathbf{1}_{P_j\neq P_Q}\langle\Delta_P f_1\rangle_{P_j}T_{P_j,Q}(x)\,d\mu(x)\Bigg|$$

$$\le\mathbf{A}_t\cdot\Bigg\|\Bigg(\sum_{Q\in\mathcal{G}_2}|\Delta_Q f_2|^2\Bigg)^{1/2}\Bigg\|_{p_2}\lesssim\mathbf{A}_t\,,$$

where we have denoted

$$\mathbf{A}_t=\sum_{j=1}^{2^n}\Bigg\|\Bigg(\sum_{k\in\mathbf{Z}}\Bigg|\sum_{P\in\mathcal{G}_{1,k+t}}\langle\Delta_P f_1\rangle_{P_j}\sum_{Q\in\mathcal{G}_{2,k}}\mathbf{1}_{Q\subset P}\mathbf{1}_{P_j\neq P_Q}\mathbf{1}_Q T_{P_j,Q}\Bigg|^2\Bigg)^{1/2}\Bigg\|_{p_1}\,.$$

By an off-diagonal estimate for $T_{P_j,Q}$, i.e. Lemma 2.11 applied to cubes P_j and Q,

$$\Big|\sum_{Q\in\mathcal{G}_{2,k}} \mathbf{1}_{Q\subset P}\mathbf{1}_{P_j\neq P_Q}\mathbf{1}_Q(x)T_{P_j,Q}(x)\Big| \lesssim 2^{-t\eta/4}\mathbf{1}_P(x)\mu(P_j)\mu(P)^{-1}, \qquad x\in\mathbf{R}^n.$$

Thus, by inequalities (2.5) and (2.4),

$$\begin{aligned}\mathbf{A}_t &\lesssim 2^{-t\eta/4}\bigg\|\bigg(\sum_{k\in\mathbf{Z}}\Big|\sum_{P\in\mathcal{G}_{1,k+t}}\langle|\Delta_P f_1|\rangle_P\mathbf{1}_P\Big|^2\bigg)^{1/2}\bigg\|_{p_1}\\ &\le 2^{-t\eta/4}\bigg\|\bigg(\sum_{k\in\mathbf{Z}}\Big(\mathbf{E}_{1,k+t}|\Delta_{k+t}f_1|\Big)^2\bigg)^{1/2}\bigg\|_{p_1} \lesssim 2^{-t\eta/4}\|f_1\|_{p_1}\lesssim 2^{-t\eta/4}.\end{aligned}$$

The last bound is summable in $t\ge r+1$. The proof of Proposition 6.1 is complete.

7. The Separated Term

Here we treat the separated term, we refer to Section 3.2.

7.1. **Proposition.** *We have inequality* $|\mathbf{B}_{\text{separated}}(f_1,f_2)|\lesssim 1$.

For the proof, we need preparations. Recall that $D(Q,P)=\ell Q+\operatorname{dist}(Q,P)+\ell P$, and write $D(Q,P)/\ell P\sim 2^u$ if $2^u<D(Q,P)/\ell P\le 2^{u+1}$. The separated term is a sum over $u,m\in\mathbb{N}_0$ and $j\in\{1,2,\ldots,2^n\}$ of terms

$$\mathbf{B}^{u,m,j}(f_1,f_2)=\sum_{k\in\mathbf{Z}}\sum_{Q\in\mathcal{G}_{2,k-m}}\sum_{\substack{P\in\mathcal{G}_{1,k}\\ D(Q,P)/\ell P\sim 2^u}}\mathbf{1}_{\ell Q\le \operatorname{dist}(Q,P)}\langle\Delta_P f_1\rangle_{P_j}\langle T\mathbf{1}_{P_j},\Delta_Q f_2\rangle.$$

For Q and P_j as in the summation above, let us write $T_{P_j,Q}=\mathbf{1}_{\ell Q\le\operatorname{dist}(Q,P)}\{T\mathbf{1}_{P_j}-T\mathbf{1}_{P_j}(x_Q)\}$. Since $\Delta_Q f_2$ has mean zero, we can write $|\mathbf{B}^{u,m,j}(f_1,f_2)|$ as

$$\begin{aligned}&\bigg|\int_{\mathbf{R}^n}\sum_{k\in\mathbf{Z}}\sum_{Q\in\mathcal{G}_{2,k-m}}\Delta_Q f_2(x)\cdot\mathbf{1}_Q(x)\sum_{\substack{P\in\mathcal{G}_{1,k}\\ D(Q,P)/\ell P\sim 2^u}}\langle\Delta_P f_1\rangle_{P_j}T_{P_j,Q}(x)\,d\mu(x)\bigg|\\ &\le \mathbf{A}_{u,m,j}\cdot\bigg\|\bigg(\sum_{k\in\mathbf{Z}}\sum_{Q\in\mathcal{G}_{2,k-m}}|\Delta_Q f_2|^2\bigg)^{1/2}\bigg\|_{p_2}\lesssim \mathbf{A}_{u,m,j},\end{aligned}$$

where we have denoted

$$\mathbf{A}_{u,m,j}=\bigg\|\bigg(\sum_{k\in\mathbf{Z}}\sum_{Q\in\mathcal{G}_{2,k-m}}\Big|\mathbf{1}_Q\sum_{\substack{P\in\mathcal{G}_{1,k}\\ D(Q,P)/\ell P\sim 2^u}}\langle\Delta_P f_1\rangle_{P_j}T_{P_j,Q}\Big|^2\bigg)^{1/2}\bigg\|_{p_1}.$$

In order to finish the proof of Proposition 7.1, we invoke the following lemma.

7.2. **Lemma.** *For* $u, m \in \mathbf{N}_0$ *and* $j \in \{1, 2, \dots, 2^n\}$*, we have* $\mathbf{A}_{u,m,j} \lesssim 2^{-\eta(m+u)/4}$.

Proof. For each $S \in \mathcal{D}_{1,k+u+\theta(u+m)}$, $k \in \mathbf{Z}$, we consider the kernel

$$K_S(x,y) = \sum_{\substack{P \in \mathcal{G}_{1,k} \\ P \subset S}} \sum_{\substack{Q \in \mathcal{G}_{2,k-m} \\ D(Q,P)/\ell P \sim 2^u}} \mathbf{1}_Q(x) \cdot \widetilde{T}_{P_j,Q}(x) \cdot \mathbf{1}_{P_j}(y)\,,$$

where $\widetilde{T}_{P_j,Q} = \mathbf{1}_{\ell Q \le \operatorname{dist}(Q,P)} \widetilde{T}_{P_j,Q}$ is defined by

$$\frac{T_{P_j,Q}(x)}{\mu(P_j)} = 2^{-\eta(m+u)/4} \cdot \frac{\widetilde{T}_{P_j,Q}(x)}{\mu(S)}\,.$$

By Lemma 2.10 and Lemma 2.11,

$$|K_S(x,y)| \lesssim \sum_{\substack{P \in \mathcal{G}_{1,k} \\ P \subset S}} \sum_{\substack{Q \in \mathcal{G}_{2,k-m} \\ D(Q,P)/\ell P \sim 2^u}} \mathbf{1}_Q(x) \cdot \mathbf{1}_{P_j}(y) \le \mathbf{1}_S(x) \cdot \mathbf{1}_S(y)\,. \tag{7.3}$$

We can now finish the proof as follows. Inequality (7.3) allows us to write $2^{\eta p_1(m+u)/4} \mathbf{A}^{p_1}_{u,m,j}$ as

$$\int_{\mathbf{R}^n} \Bigg(\sum_{k \in \mathbf{Z}} \Bigg| \sum_{S \in \mathcal{D}_{1,k+u+\theta(u+m)}} \frac{1}{\mu(S)} \int_{\mathbf{R}^n} K_S(x,y) \Delta_k f_1(y)\, d\mu(y) \Bigg|^2 \Bigg)^{p_1/2} d\mu(x)$$

$$\lesssim \int_{\mathbf{R}^n} \Bigg(\sum_{k \in \mathbf{Z}} \Bigg| \underbrace{\sum_{S \in \mathcal{D}_{1,k+u+\theta(u+m)}} \langle |\Delta_k f_1| \rangle_S \mathbf{1}_S(x)}_{= \mathbf{E}_{1,k+u+\theta(u+m)} |\Delta_k f_1|(x)} \Bigg|^2 \Bigg)^{p_1/2} d\mu(x)\,.$$

Appealing to inequalities (2.5) and (2.4) shows that $\mathbf{A}_{u,m,j} \lesssim 2^{-\eta(m+u)/4}$. □

8. Preparations for the Nearby Term

The surgery argument for the nearby term follows [19] but there are also essential differences. Let us abbreviate $(P,Q) \in \mathcal{P}_{\text{nearby}}$ as $P \sim Q$. Hence, the conditions for $P \sim Q$ are

$$(P,Q) \in \mathcal{G}_1 \times \mathcal{G}_2\,, \qquad 2^{-r}\ell P \le \ell Q \le \ell P\,, \quad \operatorname{dist}(Q,P) < \ell Q = \ell Q \wedge \ell P. \tag{8.1}$$

In particular $\ell Q \le \ell P \le D(Q,P) \le (2^r + 2)\ell Q$, i.e., these quantities are comparable if $P \sim Q$. During the course of the remaining sections, we will prove the following proposition.

8.2. **Proposition.** *For a fixed* $t > p_1 \vee p_2$*, we have*

$$\mathbf{E}_{\omega_1} \mathbf{E}_{\omega_2} |\mathbf{B}_{\text{nearby}}(f_1, f_2)| \le C(r, \upsilon, \epsilon)(1 + \mathbf{T}_{\text{loc}}) + \Big(C(r,\upsilon)\epsilon^{1/t} + C(r)\upsilon^{1/t} \Big) \mathbf{T}\,.$$

Aside from the indicated absorption parameters, the constants on the right hand side can depend upon the parameters n, p_1, p_2, η, μ.

8.3. *Remark.* We shall track dependence of various inequalities on absorption parameters: r, υ, ϵ. There is no need to do this quantitatively, and thus we agree upon the following convenient notation: $C(r)$, $C(r,\upsilon)$, and $C(r,\upsilon,\epsilon)$ denote positive numbers that are allowed to depend on the indicated absorption parameters, but also on parameters n, p_1, p_2, η, μ. Moreover, the value of these numbers is allowed to vary from one occurrence to another.

For a given $P \in \mathcal{G}_1$ there are at most $C(r)$ cubes $Q \in \mathcal{G}_2$ satisfying (8.1). Hence, without loss of generality, it suffices consider a finite number of subseries of the general form

$$\mathbf{E}_{\omega_1}\mathbf{E}_{\omega_2}\Big| \sum_{P\in\mathcal{G}_1} \langle T\Delta_P f_1, \Delta_Q f_2\rangle \Big|, \tag{8.4}$$

where $Q = Q(P) \in \mathcal{G}_2 \cup \{\emptyset\}$ inside the summation satisfies $P \sim Q$ or $Q = \emptyset$.[2] At the same time, we can also assume that for any $Q \in \mathcal{G}_2$ there is at most one $P \in \mathcal{G}_1$ such that $Q = P(Q)$. We fix one series like this, and the convention that Q is implicitly a function of P will be maintained.

8.1. **First reductions.** We immediately find that (8.4) is dominated by

$$\sum_{i,j=1}^{2^n} \mathbf{E}_{\omega_1}\mathbf{E}_{\omega_2}\Big| \sum_{P\in\mathcal{G}_1} \langle \Delta_P f_1\rangle_{P_j} \langle T\mathbf{1}_{P_j}, \mathbf{1}_{Q_i}\rangle \langle \Delta_Q f_2\rangle_{Q_i} \Big|. \tag{8.5}$$

Fix $i,j \in \{1,\dots,n\}$. For a cube R in $\mathbf{R}^n$, define an 'υ-boundary region': $\delta_R^\upsilon = (1+\upsilon)R \setminus (1-\upsilon)R$. If $P \in \mathcal{D}_1$ and $Q = Q(P) \neq \emptyset$, we write

$$\begin{aligned} &Q_{i,\partial} = Q_i \cap \delta_{P_j}^\upsilon; \quad Q_{i,\mathrm{sep}} = (Q_i \setminus Q_{i,\partial}) \setminus (Q_i \cap P_j); \quad \Delta_{Q_i} = (Q_i \cap P_j) \setminus (Q_{i,\partial}); \\ &P_{j,\partial} = P_j \cap \delta_{Q_i}^\upsilon; \quad P_{j,\mathrm{sep}} = (P_j \setminus P_{j,\partial}) \setminus (Q_i \cap P_j); \quad \Delta_{P_j} = (Q_i \cap P_j) \setminus P_{j,\partial}. \end{aligned} \tag{8.6}$$

For an illustration of these sets, we refer to Figure 1.

We write the matrix coefficient $\langle T\mathbf{1}_{P_j}, \mathbf{1}_{Q_i}\rangle$ in (8.5) as

$$\langle T\mathbf{1}_{P_{j,\mathrm{sep}}}, \mathbf{1}_{Q_i}\rangle + \langle T\mathbf{1}_{P_{j,\partial}}, \mathbf{1}_{Q_i}\rangle + \langle T\mathbf{1}_{\Delta_{P_j}}, \mathbf{1}_{\Delta_{Q_i}}\rangle + \langle T\mathbf{1}_{\Delta_{P_j}}, \mathbf{1}_{Q_{i,\partial}}\rangle + \langle T\mathbf{1}_{\Delta_{P_j}}, \mathbf{1}_{Q_{i,\mathrm{sep}}}\rangle, \tag{8.7}$$

and these are denoted by $M_1(P) + M_2(P) + M_3(P) + M_4(P) + M_5(P)$, respectively.

8.2. **Description of different terms.** The heart of the argument lies in estimating terms

$$M_3(P) = \langle \mathbf{1}_{\Delta_{P_j}}, T\mathbf{1}_{\Delta_{Q_i}}\rangle = \alpha_1(P) + \alpha_2(P) + \alpha_3(P),$$

where the last decomposition depends on a third random dyadic system $\mathcal{D}_3$, we refer to (8.9). Terms $\alpha_2(P)$ and $\alpha_3(P)$, along with $M_2(P)$ and $M_4(P)$, are 'υ-boundary' terms. The 'separated' terms $M_1(P)$ and $M_5(P)$ are treated by kernel size condition.

[2]We agree that $\Delta_\emptyset f_2 = 0$.

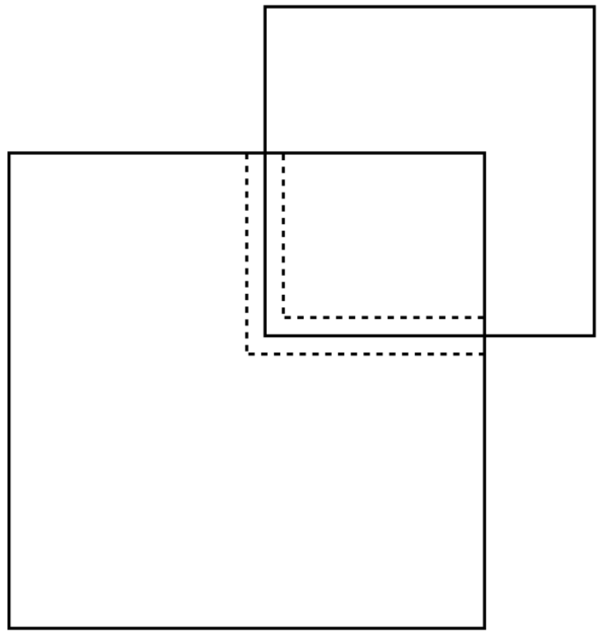

FIGURE 1. The larger cube is P_j, and the smaller cube is Q_i. The dashed line segments separate sets $P_{j,sep}$, $P_{j,\partial}$, and Δ_{P_j} from each other.

The term $\alpha_1(P)$ will further be expanded in (8.10) as

$$\alpha_1(P) = \beta_1(P) + \beta_2(P) + \beta_3(P),$$

where $\beta_1(P)$ and $\beta_2(P)$ are so called 'ϵ-boundary' terms. The local testing conditions and kernel size estimates are exploited in estimating 'intersecting' term $\beta_3(P)$.

8.3. **Decomposition of $M_3(P)$.** Without loss of generality, we can assume that $\Delta_{Q_i} \neq \emptyset$ and $\Delta_{P_j} \neq \emptyset$. Indeed, otherwise we already have $M_3(P) = 0$.

We introduce a third random dyadic system $\mathcal{D}_3 = \mathcal{D}(\omega_3)$ that is independent of both $\mathcal{D}_1$ and $\mathcal{D}_2$. Fix $j(\upsilon) \in Z$ such that $\upsilon/64 \leq 2^{j(\upsilon)} < \upsilon/32$. Then, for every $P \in \mathcal{G}_1$ with $Q = Q(P) \neq \emptyset$, we define a layer

$$\mathcal{L} = \mathcal{L}(P, \upsilon) := \mathcal{D}_{3,\log_2(s)}$$

of $\mathcal{D}_3$-cubes with side length

$$s = 2^{j(\upsilon)}\ell Q_i = 2^{j(\upsilon)} \cdot (\ell Q_i \wedge \ell P_j)\,. \tag{8.8}$$

That is, $\mathcal{L}$ is a layer of $\mathcal{D}_3$ that depends on parameters P and υ.

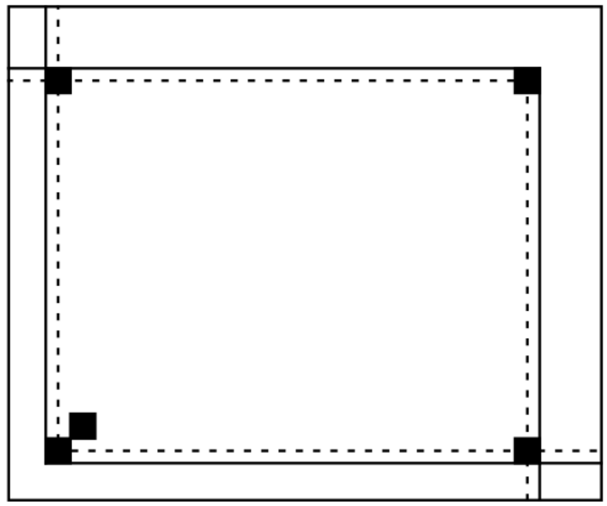

FIGURE 2. Parallelogram is $Q_i \cap P_j$. Interiors of some black $\mathcal{L}$-cubes intersect dashed line segments, which belong to the boundaries of either Δ_{P_j} or Δ_{Q_i}. The $\mathcal{L}$-adjusted sets $\Delta_{Q_i}^{\mathcal{L}}$ and $\Delta_{P_j}^{\mathcal{L}}$, with solid boundaries, do not intersect the indicated $\mathcal{L}$-cubes.

Let $\Delta_{Q_i}^{\mathcal{L}}, \Delta_{P_j}^{\mathcal{L}} \subset Q_i \cap P_j$ be the following adaptations of Δ_{Q_i} and Δ_{P_j} to $\mathcal{L}$. If necessary, we enlargen the latter sets so that, for every $G \in \mathcal{L}$, either $G \cap \Delta_{Q_i}^{\mathcal{L}} = G \cap \Delta_{P_j}^{\mathcal{L}} = G$ or one of the two intersections $G \cap \Delta_{Q_i}^{\mathcal{L}}$ and $G \cap \Delta_{P_j}^{\mathcal{L}}$ is empty. This is done in such a way that we can write

$$\Delta_{Q_i}^{\mathcal{L}} = \Delta_{Q_i} \cup \Delta_{Q_i}^{\partial}, \quad \Delta_{P_j}^{\mathcal{L}} = \Delta_{P_j} \cup \Delta_{P_j}^{\partial},$$

both as disjoint unions, such that $\Delta_{Q_i}^{\partial} \subset Q_{i,\partial} \cap P_j$ and $\Delta_{P_j}^{\partial} \subset P_{j,\partial} \cap Q_i$. For an illustration, we refer to Figure 2.

Now observe that $M_3(P) = \langle T\mathbf{1}_{\Delta_{P_j}}, \mathbf{1}_{\Delta_{Q_i}} \rangle$ can be written as

$$\alpha_1(P) + \alpha_2(P) + \alpha_3(P) = \langle T\mathbf{1}_{\Delta_{P_j}^{\mathcal{L}}}, \mathbf{1}_{\Delta_{Q_i}^{\mathcal{L}}} \rangle - \langle T\mathbf{1}_{\Delta_{P_j}^{\partial}}, \mathbf{1}_{\Delta_{Q_i}^{\mathcal{L}}} \rangle - \langle T\mathbf{1}_{\Delta_{P_j}}, \mathbf{1}_{\Delta_{Q_i}^{\partial}} \rangle . \tag{8.9}$$

We remark that the terms in this decomposition depends on $\mathcal{D}_3$.

In order to define ϵ-boundary terms, we let $P \in \mathcal{G}_1$ and write

$$L_\epsilon = L_\epsilon(P, \upsilon) = \bigcup_{G \in \mathcal{L}(P,\upsilon)} \delta_G^\epsilon, \quad \delta_G^\epsilon = (1+\epsilon)G \setminus (1-\epsilon)G.$$

We also write $\widetilde{G} = G \setminus L_\epsilon$ if $G \in \mathcal{L} = \mathcal{L}(P, \upsilon)$. Define

$$\Delta'_{Q_i} = \Delta_{Q_i}^{\mathcal{L}} \cap L_\epsilon, \quad \widetilde{\Delta}_{Q_i} = \Delta_{Q_i}^{\mathcal{L}} \setminus L_\epsilon, \quad \Delta'_{P_j} = \Delta_{P_j}^{\mathcal{L}} \cap L_\epsilon, \quad \widetilde{\Delta}_{P_j} = \Delta_{P_j}^{\mathcal{L}} \setminus L_\epsilon .$$

Finally, we write $\alpha_1(P) = \langle T\mathbf{1}_{\Delta^{\mathcal{L}}_{P_j}}, \mathbf{1}_{\Delta^{\mathcal{L}}_{Q_i}}\rangle$ as

$$\beta_1(P) + \beta_2(P) + \beta_3(P) = \langle T\mathbf{1}_{\Delta'_{P_j}}, \mathbf{1}_{\Delta^{\mathcal{L}}_{Q_i}}\rangle + \langle T\mathbf{1}_{\widetilde{\Delta}_{P_j}}, \mathbf{1}_{\Delta'_{Q_i}}\rangle + \langle T\mathbf{1}_{\widetilde{\Delta}_{P_j}}, \mathbf{1}_{\widetilde{\Delta}_{Q_i}}\rangle\,. \tag{8.10}$$

9. The Nearby-Non-Boundary Term

We estimate summations involving the separated terms $M_1(P)$ and $M_5(P)$, and the intersecting term $\beta_3(P)$. All of the estimates will be uniform over all three dyadic grids.

9.1. **Separated term.** The two indicators appearing in either $M_1(P)$ or $M_5(P)$ are associated with sets separated from each other. This observation will allow us to prove inequality

$$\Big|\sum_{P\in\mathcal{G}_1} \langle \Delta_P f_1\rangle_{P_j}(M_1(P) + M_5(P))\langle \Delta_Q f_2\rangle_{Q_i}\Big| \le C(r,\upsilon)\,. \tag{9.1}$$

Proof of inequality (9.1). We focus on summation over the terms $M_1(P)$, and the treatment of summation over terms $M_5(P)$ is analogous. We write $T_{P_j,Q_i} = \mathbf{1}_{Q=Q(P)}\langle T\mathbf{1}_{P_{j,\mathrm{sep}}}, \mathbf{1}_{Q_i}\rangle$. Then, by inequalities (8.1), the term under focus can be written as

$$\begin{aligned}
&\Big|\sum_{m=0}^{r}\sum_{u\in\{0,1\}}\sum_{k\in\mathbf{Z}}\sum_{Q\in\mathcal{G}_{2,k-m}}\sum_{\substack{P\in\mathcal{G}_{1,k}\\ D(Q,P)/\ell P\sim 2^u}} \langle\Delta_P f_1\rangle_{P_j} T_{P_j,Q_i}\langle\Delta_Q f_2\rangle_{Q_i}\Big| \\
&\le \sum_{m,u}\mathbf{A}_{m,u,i,j}\cdot\Big\|\Big(\sum_{k\in\mathbf{Z}}\sum_{Q\in\mathcal{G}_{2,k-m}}|\Delta_Q f_2|^2\Big)^{1/2}\Big\|_{p_2} \lesssim \sum_{m,u}\mathbf{A}_{m,u,i,j}\,,
\end{aligned}$$

where we have denoted

$$\mathbf{A}_{m,u,i,j} = \Big\|\Big(\sum_{k\in\mathbf{Z}}\sum_{Q\in\mathcal{G}_{2,k-m}}\Big|\mathbf{1}_{Q_i}\sum_{\substack{P\in\mathcal{G}_{1,k}\\ D(Q,P)/\ell P\sim 2^u}}\langle\Delta_P f_1\rangle_{P_j}\frac{T_{P_j,Q_i}}{\mu(Q_i)}\Big|^2\Big)^{1/2}\Big\|_{p_1}\,.$$

The proof of inequality (9.1) is finished by invoking Lemma 9.2 below. □

9.2. **Lemma.** *For $m\in\{0,1,\dots,r\}$ and $u\in\{0,1\}$, we have $|\mathbf{A}_{m,u,i,j}|\le C(r,\upsilon)$.*

Proof. For each $k\in\mathbf{Z}$ and $S\in\mathcal{D}_{1,k+u+\theta(u+m)}$, define a kernel

$$K_S(x,y) = \sum_{\substack{P\in\mathcal{G}_{1,k}\\ P\subset S}}\sum_{\substack{Q\in\mathcal{G}_{2,k-m}\\ D(Q,P)/\ell P\sim 2^u}}\mathbf{1}_{Q_i}(x)\cdot\widetilde{T}_{P_j,Q_i}\cdot\mathbf{1}_{P_j}(y)\,,\qquad x,y\in\mathbf{R}^n\,,$$

where $\widetilde{T}_{P_j,Q_i} = \mathbf{1}_{Q=Q(P)}\widetilde{T}_{P_j,Q_i}$ is defined by

$$\frac{T_{P_j,Q_i}}{\mu(P_j)\mu(Q_i)} = \frac{\widetilde{T}_{P_j,Q_i}}{\mu(S)}\,.$$

Consider cubes P and Q as in the definition of K_S, and let $y \in P_{j,\mathrm{sep}}$ and $x \in Q_i$. By the upper doubling properties of μ, and the facts that $|x-y| \geq \upsilon 2^{-r}\ell P_j$ and $S \subset B(y, 2^{1+u+\theta(u+m)}\ell P)$, we find that $\lambda(y,|x-y|)^{-1} \leq C(r,\upsilon)\mu(S)^{-1}$. Hence, by definition,

$$|T_{P_j,Q_i}| \leq \int_{Q_i}\int_{P_{j,\mathrm{sep}}} \frac{1}{\lambda(y,|x-y|)}\,d\mu(y)\,d\mu(x) \leq C(r,\upsilon)\mu(Q_i)\mu(P_j)\mu(S)^{-1}\,.$$

As a consequence $|\widetilde{T}_{P_j,Q_i}| \leq C(r,\upsilon)$ and, by recalling Lemma 2.10,

$$|K_S(x,y)| \leq C(r,\upsilon)\sum_{\substack{P\in\mathcal{G}_{1,k}\\ P\subset S}}\ \sum_{\substack{P\in\mathcal{G}_{2,k-m}\\ D(Q,P)/\ell P\sim 2^u}} \mathbf{1}_{Q_i}(x)\cdot\mathbf{1}_{P_j}(y) \leq C(r,\upsilon)\cdot\mathbf{1}_S(x)\cdot\mathbf{1}_S(y)\,.$$

After these preparations, we finish the proof by proceeding as in Lemma 7.2. □

9.2. **Intersecting term.** The following inequality deals with intersecting part, i.e., terms $\beta_3(P)$;

$$\left|\sum_{P\in\mathcal{G}_1}\langle\Delta_P f_1\rangle_{P_j}\beta_3(P)\langle\Delta_Q f_2\rangle_{Q_i}\right| \leq C(r,\upsilon,\epsilon)(1+\mathbf{T}_{\mathrm{loc}})\,. \tag{9.3}$$

The proof of this inequality relies on the kernel size estimate and local testing conditions.

Proof of inequality (9.3). We tacitly restrict all the summations here to cubes $P \in \mathcal{G}_1$ for which $\mu(Q_i\cap P_j) \neq 0$. Indeed, otherwise $\beta_3(P) = 0$. By writing $\mu(Q_i\cap P_j) = \int \mathbf{1}_{Q_i}\mathbf{1}_{P_j}\,d\mu$ and using Cauchy-Schwarz and Hölder's inequality,

$$\begin{aligned}&\left|\sum_{P\in\mathcal{G}_1}\langle\Delta_P f_1\rangle_{P_j}\beta_3(P)\langle\Delta_Q f_2\rangle_{Q_i}\right|\\ &\leq \left\|\left(\sum_{P\in\mathcal{G}_1}|\langle\Delta_P f_1\rangle_{P_j}\mathbf{1}_{P_j}|^2\right)^{1/2}\right\|_{p_1}\cdot\left\|\left(\sum_{P\in\mathcal{G}_1}\left|\frac{\beta_3(P)}{\mu(Q_i\cap P_j)}\langle\Delta_Q f_2\rangle_{Q_i}\mathbf{1}_{Q_i}\right|^2\right)^{1/2}\right\|_{p_2}.\end{aligned}$$

By inequality (2.4), the first factor is bounded by $\lesssim 1$. Let us then focus on the second factor; by writing the summation in terms of Q and using Lemma (9.4), we obtain an upper bound $C(r,\upsilon,\epsilon)(1+\mathbf{T}_{\mathrm{loc}})$ for the second term. □

9.4. **Lemma.** *Let* $P\in\mathcal{G}_1$. *Then* $|\beta_3(P)| \leq C(r,\upsilon,\epsilon)(1+\mathbf{T}_{\mathrm{loc}})\mu(Q_i\cap P_j)$.

Proof. We can assume that $Q = Q(P) \neq \emptyset$, hence $P\sim Q$. Consider the expansion,

$$\beta_3(P) = \langle T\mathbf{1}_{\widetilde{\Delta}_{P_j}},\mathbf{1}_{\widetilde{\Delta}_{Q_i}}\rangle = \sum_{\substack{G,H\in\mathcal{L}\\ G\neq H}}\langle T(\mathbf{1}_G\mathbf{1}_{\widetilde{\Delta}_{P_j}}),\mathbf{1}_H\mathbf{1}_{\widetilde{\Delta}_{Q_i}}\rangle + \sum_{G\in\mathcal{L}}\langle T(\mathbf{1}_G\mathbf{1}_{\widetilde{\Delta}_{P_j}}),\mathbf{1}_G\mathbf{1}_{\widetilde{\Delta}_{Q_i}}\rangle.$$

In both of the series above, the finite number of summands depends on n and υ. Hence, it suffices to obtain estimates for individual summands for fixed $G, H \in \mathcal{L}$. First, if $G \neq H$, then

$$\ell Q_i \le C(r,\upsilon,\epsilon)\,\mathrm{dist}(G \cap \widetilde{\Delta}_{P_j}, H \cap \widetilde{\Delta}_{Q_i}).$$

In particular, $\lambda(x,|x-y|)^{-1} \le C(r,\upsilon,\epsilon)\mu(Q_i)^{-1}$ if $x \in H \cap \widetilde{\Delta}_{Q_i}$ and $y \in G \cap \widetilde{\Delta}_{P_j}$. Hence,

$$\begin{aligned} |\langle T(\mathbf{1}_G \mathbf{1}_{\widetilde{\Delta}_{P_j}}), \mathbf{1}_H \mathbf{1}_{\widetilde{\Delta}_{Q_i}}\rangle| &\le \int_{H\cap\widetilde{\Delta}_{Q_i}} \int_{G\cap\widetilde{\Delta}_{P_j}} \frac{1}{\lambda(x,|x-y|)}\, d\mu(y)\, d\mu(x) \\ &\le C(r,\upsilon,\epsilon)\frac{\mu(Q_i\cap P_j)\mu(Q_i\cap P_j)}{\mu(Q_i)} \le C(r,\upsilon,\epsilon)\mu(Q_i \cap P_j). \end{aligned}$$

In the last step, we also used the fact that $\widetilde{\Delta}_{P_j} \cup \widetilde{\Delta}_{Q_i} \subset Q_i \cap P_j$.

Then we consider the case of $G = H$. By construction,

$$\langle T(\mathbf{1}_G \mathbf{1}_{\widetilde{\Delta}_{P_j}}), \mathbf{1}_G \mathbf{1}_{\widetilde{\Delta}_{Q_i}}\rangle = \begin{cases} \langle T\mathbf{1}_{\widetilde{G}}, \mathbf{1}_{\widetilde{G}}\rangle, & \text{if } G = G\cap \Delta^{\mathcal{L}}_{P_j} = G \cap \Delta^{\mathcal{L}}_{Q_i}\,; \\ 0 & \text{otherwise.} \end{cases}$$

In any case, by local testing conditions $|\langle T(\mathbf{1}_G \mathbf{1}_{\widetilde{\Delta}_{P_j}}), \mathbf{1}_G \mathbf{1}_{\widetilde{\Delta}_{Q_i}}\rangle| \le \mathbf{T}_{\mathrm{loc}}\mu(\widetilde{G}) \le \mathbf{T}_{\mathrm{loc}}\mu(Q_i \cap P_j)$. □

10. The Nearby-Boundary Term

Here we treat the ϵ and υ boundary terms by probabilistic arguments.

10.1. **The ϵ-boundary terms.** Following inequality controls summation for ϵ-boundary terms. Let $t > p_1 \vee p_2$ be a positive real number. Then

$$\mathbf{E}_{\omega_3}\Big| \sum_{P\in\mathcal{G}_1} \langle \Delta_P f_1\rangle_{P_j} \big(\beta_1(P) + \beta_2(P)\big) \langle \Delta_Q f_2\rangle_{Q_i}\Big| \le C(r,\upsilon)\epsilon^{1/t}\mathbf{T}\,. \tag{10.1}$$

The expectations over the dyadic system $\mathcal{D}_3$ are crucial here, and here only.

We let $\epsilon = (\varepsilon_k)_{k\in\mathbf{Z}}$ be a sequence of Rademacher variables, supported on a probability space $(\Omega, \mathbf{P})$. We can also associate Rademacher variables to $\mathcal{D}_j$-dyadic cubes with $j \in \{1,2\}$:— fix an injection $R \mapsto j(R) : \mathcal{D}_j \to \mathbf{Z}$, and use notation $\varepsilon_R = \varepsilon_{j(R)}$.

We rely on the following improvement of the contraction principle, [18, Lemma 3.1].

10.2. **Proposition.** *Suppose that $\{\rho_k : k \in \mathbf{Z}\} \subset L^t(\widetilde{\Omega})$ for some σ-finite measure space $(\widetilde{\Omega}, \widetilde{\mathbf{P}})$ and $t \in (2,\infty)$. Then, for all complex-valued sequences $(\xi_k)_{k\in\mathbf{Z}}$,*

$$\Big\| \sum_{k=-\infty}^{\infty} \varepsilon_k \rho_k \xi_k \Big\|_{L^t(\widetilde{\Omega};L^2(\Omega))} \lesssim \sup_{k\in\mathbf{Z}} \|\rho_k\|_{L^t(\widetilde{\Omega})} \cdot \Big\| \sum_{k=-\infty}^{\infty} \varepsilon_k \xi_k \Big\|_{L^2(\Omega)}.$$

Proof of inequality (10.1). Let us focus on the sum involving the terms $\beta_1(P)$; the estimate for the sum involving terms $\beta_2(P)$ is similar. We randomize and use Hölder's inequality,

$$\begin{aligned}
&\Big|\sum_{P\in\mathcal{G}_1}\langle\Delta_P f_1\rangle_{P_j}\langle T\mathbf{1}_{\Delta'_{P_j}},\mathbf{1}_{\Delta^{\mathcal{L}}_{Q_i}}\rangle\langle\Delta_Q f_2\rangle_{Q_i}\Big| \\
&= \Big|\int_\Omega \Big\langle T\Big(\sum_{S\in\mathcal{G}_1}\varepsilon_S\mathbf{1}_{\Delta'_{S_j}}\langle\Delta_S f_1\rangle_{S_j}\Big),\sum_{P\in\mathcal{G}_1}\varepsilon_P\mathbf{1}_{\Delta^{\mathcal{L}}_{Q_i}}\langle\Delta_Q f_2\rangle_{Q_i}\Big\rangle d\mathbf{P}(\epsilon)\Big| \qquad (10.3)\\
&\le \Big\|T\Big(\sum_{S\in\mathcal{G}_1}\varepsilon_S\mathbf{1}_{\Delta'_{S_j}}\langle\Delta_S f_1\rangle_{S_j}\Big)\Big\|_{L^{p_1}(\Omega\times\mathbf{R}^n)}\Big\|\sum_{P\in\mathcal{G}_1}\varepsilon_P\mathbf{1}_{\Delta^{\mathcal{L}}_{Q_i}}\langle\Delta_Q f_2\rangle_{Q_i}\Big\|_{L^{p_2}(\Omega\times\mathbf{R}^n)}.
\end{aligned}$$

Index the very last summation in terms of $\mathcal{D}_2$. This can be done by using our standing assumptions of $P\mapsto Q(P)=Q$. Then, by the contraction principle and inequality $|\mathbf{1}_{\Delta^{\mathcal{L}}_{Q_i}}|\le\mathbf{1}_{Q_i}$, we see that the second factor in the last line of (10.3) is bounded (up to a constant multiple) by $\|f_2\|_{p_2}\lesssim 1$.

In order to estimate the first factor in the last line of (10.3) we first extract operator norm $\mathbf{T}$. Then we fix $S\in\mathcal{G}_{1,k}$ with $k\in Z$. By (8.1) and (8.8),

$$\Delta'_{S_j}\subset L_\epsilon(S,\upsilon)=\bigcup_{G\in\mathcal{L}(S,\upsilon)}\delta^\epsilon_G\subset\bigcup_{m=j(\upsilon)+k-r-1}^{j(\upsilon)+k-1}\ \bigcup_{G\in\mathcal{D}_{3,m}}\delta^\epsilon_G=:\delta^\epsilon(k).$$

Hence, we have $\mathbf{1}_{\Delta'_{S_j}}\le\mathbf{1}_{\delta^\epsilon(k)}\mathbf{1}_{S_j}$. By the contraction principle and assumption $t\ge p_1$,

$$\begin{aligned}
&\mathbf{E}_{\omega_3}\Big\|\sum_{S\in\mathcal{G}_1}\varepsilon_S\mathbf{1}_{\Delta'_{S_j}}\langle\Delta_S f_1\rangle_{S_j}\Big\|_{L^{p_1}(\Omega\times\mathbf{R}^n)}\lesssim\mathbf{E}_{\omega_3}\Big\|\sum_{k\in Z}\varepsilon_k\mathbf{1}_{\delta^\epsilon(k)}\sum_{S\in\mathcal{D}_{1,k}}\mathbf{1}_{S_j}\langle\Delta_S f_1\rangle_{S_j}\Big\|_{L^{p_1}(\Omega\times\mathbf{R}^n)} \qquad (10.4)\\
&\le\Bigg(\int_{\mathbf{R}^n}\Bigg[\mathbf{E}_{\omega_3}\Big\|\sum_{k\in Z}\varepsilon_k\mathbf{1}_{\delta^\epsilon(k)}(x)\sum_{S\in\mathcal{D}_{1,k}}\mathbf{1}_{S_j}(x)\langle\Delta_S f_1\rangle_{S_j}\Big\|^t_{L^{p_1}(\Omega)}\Bigg]^{p_1/t}d\mu(x)\Bigg)^{1/p_1}.
\end{aligned}$$

For a fixed $x\in\mathbf{R}^n$, the last integrand evaluated at x is of the form as in Proposition 10.2 with $\xi_k=\sum_{S\in\mathcal{D}_{1,k}}\mathbf{1}_{S_j}(x)\langle\Delta_S f_1\rangle_{S_j}$. Moreover, the random variables $\rho_k:=\mathbf{1}_{\delta^\epsilon(k)}(x)$ as functions of $\omega_3\in\Omega_3=(\{0,1\}^n)^{\mathbf{Z}}$ belong to $L^t(\Omega_3)$, and they satisfy

$$\sup_{k\in Z}\|\mathbf{1}_{\delta^\epsilon(k)}(x)\|_{L^t(\Omega_3)}=\sup_{k\in Z}\mathbf{P}_{\omega_3}(\mathbf{1}_{\delta^\epsilon(k)}(x)=1)^{1/t}\le C(r,\upsilon)\epsilon^{1/t}.$$

Hence, by Proposition 10.2 and Khintchine's inequality,

$$\mathrm{LHS}(10.4)\le C(r,\upsilon)\epsilon^{1/t}\Big\|\sum_{S\in\mathcal{D}_1}\varepsilon_S\mathbf{1}_{S_j}\langle\Delta_S f_1\rangle_{S_j}\Big\|_{L^{p_1}(\Omega\times\mathbf{R}^n)}\le C(r,\upsilon)\epsilon^{1/t}.$$

The proof is complete. □

10.2. **The υ-boundary terms.** The following inequality controls summation of the υ-boundary terms. Let $t > p_1 \vee p_2$. Then

$$\mathbf{E}_{\omega_1}\mathbf{E}_{\omega_2}\Big|\sum_{P\in\mathcal{G}_1}\langle\Delta_P f_1\rangle_{P_j}(M_2(P)+M_4(P)+\alpha_2(P)+\alpha_3(P))\langle\Delta_Q f_2\rangle_{Q_i}\Big| \le C(r)\upsilon^{1/t}\mathbf{T}\,. \tag{10.5}$$

Before the proof, let us remark that although both $\alpha_2(P)$ and $\alpha_3(P)$ depend on the random dyadic system $\mathcal{D}_3$, the inequality is uniform over all such systems.

Proof of inequality (10.5). First we observe that functions f_j depend on *both* dyadic systems, as they are (essentially) projections to good cubes. This dependency is not allowed in the argument below. Fortunately, this issue can be easily addressed—if $Q(P) \neq \emptyset$ in the series above, we have both $P \in \mathcal{G}_1$ and $Q = Q(P) \in \mathcal{G}_2$. Then, in particular $\Delta_P f_1 = \Delta_P \widetilde{f}_1$ and $\Delta_Q f_2 = \Delta_Q \widetilde{f}_2$. Functions $\widetilde{f}_j$ do not depend on the dyadic systems, and we use them to replace f_j's.

By (8.7) and (8.9), $M_2(P)+\alpha_2(P)$ and $M_4(P)+\alpha_3(P)$ are given by

$$\langle T\mathbf{1}_{P_{j,\partial}},\mathbf{1}_{Q_i}\rangle - \langle T\mathbf{1}_{\Delta^{\partial}_{P_j}},\mathbf{1}_{\Delta^{\mathcal{L}}_{Q_i}}\rangle;\quad \langle T\mathbf{1}_{\Delta_{P_j}},\mathbf{1}_{Q_{i,\partial}}\rangle - \langle T\mathbf{1}_{\Delta_{P_j}},\mathbf{1}_{\Delta^{\partial}_{Q_i}}\rangle\,,$$

respectively. Observe that

$$(\mathbf{1}_{P_{j,\partial}}+\mathbf{1}_{\Delta^{\partial}_{P_j}}) \lesssim \mathbf{1}_{P_{j,\partial}},\quad (\mathbf{1}_{Q_i}+\mathbf{1}_{\Delta^{\mathcal{L}}_{Q_i}}) \lesssim \mathbf{1}_{Q_i}\,,\quad \mathbf{1}_{\Delta_{P_j}} \lesssim \mathbf{1}_{P_j},\quad (\mathbf{1}_{Q_{i,\partial}}+\mathbf{1}_{\Delta^{\partial}_{Q_i}}) \lesssim \mathbf{1}_{Q_{i,\partial}}. \tag{10.6}$$

pointwise μ-almost everywhere. By triangle inequality, it suffices to estimate the following sums: one involving terms $m(P) \in \{M_2(P),\alpha_2(P)\}$, and the other involving terms in $\{M_4(P),\alpha_3(P)\}$. We focus on the first sum; the second one is estimated in an analogous manner, using $\mathbf{E}_{\omega_1}$.

By randomizing, using Hölder's inequality, extracting the operator norm of T, and applying the contraction principle with inequalities (10.6),

$$\begin{aligned}&\mathbf{E}_{\omega_2}\Big|\sum_{P\in\mathcal{G}_1}\langle\Delta_P\widetilde{f}_1\rangle_{P_j}m(P)\langle\Delta_Q\widetilde{f}_2\rangle_{Q_i}\Big|\\ &\lesssim \mathbf{T}\cdot\mathbf{E}_{\omega_2}\Big\{\Big\|\sum_{S\in\mathcal{G}_1}\varepsilon_S\mathbf{1}_{S_{j,\partial}}\langle\Delta_P\widetilde{f}_1\rangle_{S_j}\Big\|_{L^{p_1}(\Omega\times\mathbf{R}^n)}\Big\|\sum_{Q\in\mathcal{D}_2}\varepsilon_Q\mathbf{1}_{Q_i}\langle\Delta_Q\widetilde{f}_2\rangle_{Q_i}\Big\|_{L^{p_2}(\Omega\times\mathbf{R}^n)}\Big\}.\end{aligned}$$

By the contraction principle, we find that the last factor is ω_2-uniformly bounded by $\lesssim \|\widetilde{f}_2\|_{p_2} = 1$. In order to treat the remaining factor, we write

$$\delta^{\upsilon}(k) = \bigcup_{m=k-r-1}^{k-1}\bigcup_{Q\in\mathcal{D}_{2,m}}\delta^{\upsilon}_Q.$$

By (8.1) and (8.6), $\mathbf{1}_{S_{j,\partial}} \leq \mathbf{1}_{S_j}\mathbf{1}_{\delta^{\upsilon}_{Q_i}} \leq \mathbf{1}_{S_j}\mathbf{1}_{\delta^{\upsilon}(k)}$ if $Q = Q(S)$ with $S \in \mathcal{G}_{1,k}$. Fix $x \in \mathbf{R}^n$. The random variables $\rho_k := \mathbf{1}_{\delta^{\upsilon}(k)}(x)$ as functions of $\omega_2 \in (\{0,1\}^n)^{\mathbb{Z}}$ belong to $L^t((\{0,1\}^n)^{\mathbb{Z}})$,

$$\sup_{k\in\mathbb{Z}} \|\mathbf{1}_{\delta^{\upsilon}(k)}(x)\|_{L^t((\{0,1\}^n)^{\mathbb{Z}})} = \sup_{k\in\mathbb{Z}} \mathbf{P}_{\omega_2}(\mathbf{1}_{\delta^{\upsilon}(k)}(x) = 1)^{1/t} \leq C(r)\upsilon^{1/t}.$$

Hence, proceeding as in connection with (10.4), we find that

$$\mathbf{E}_{\omega_2}\Big\| \sum_{S\in\mathcal{G}_1} \varepsilon_S \mathbf{1}_{S_{j,\partial}} \langle \Delta_P \widetilde{f}_1\rangle_{S_j} \Big\|_{L^{p_1}(\Omega\times\mathbf{R}^n)} \leq C(r)\upsilon^{1/t} \Big\| \sum_{P\in\mathcal{D}_1} \varepsilon_P \mathbf{1}_{P_j} \langle \Delta_P \widetilde{f}_1\rangle_{P_j} \Big\|_{L^{p_1}(\Omega\times\mathbf{R}^n)}.$$

The last term is bounded by a constant multiple of $C(r)\upsilon^{1/t}$. □

References

[1] P. Auscher, S. Hofmann, C. Muscalu, T. Tao, and C. Thiele, *Carleson measures, trees, extrapolation, and* $T(b)$ *theorems*, Publ. Mat. **46** (2002), no. 2, 257–325.

[2] Pascal Auscher and Eddy Routin, *Local Tb Theorems and Hardy Inequalities*, J. Geom. Anal. **23** (2013), no. 1, 303–374.

[3] Pascal Auscher and Qi Xiang Yang, *BCR algorithm and the* $T(b)$ *theorem*, Publ. Mat. **53** (2009), no. 1, 179–196.

[4] Jean Bourgain, *Vector-valued singular integrals and the* H^1*-BMO duality*, Probability theory and harmonic analysis (Cleveland, Ohio, 1983), Monogr. Textbooks Pure Appl. Math., vol. 98, Dekker, New York, 1986, pp. 1–19.

[5] D. L. Burkholder, *Boundary value problems and sharp inequalities for martingale transforms*, Ann. Probab. **12** (1984), no. 3, 647–702.

[6] Michael Christ, *A* $T(b)$ *theorem with remarks on analytic capacity and the Cauchy integral*, Colloq. Math. **60/61** (1990), no. 2, 601–628.

[7] Guy David and Jean-Lin Journé, *A boundedness criterion for generalized Calderón-Zygmund operators*, Ann. of Math. (2) **120** (1984), no. 2, 371–397.

[8] Steve Hofmann, *A proof of the local* Tb *Theorem for standard Calderón-Zygmund operators* (2007), available at http://arxiv.org/abs/0705.0840.

[9] Guoen Hu, Yan Meng, and Dachun Yang, *A new characterization of regularized BMO spaces on non-homogeneous spaces and its applications*, Ann. Acad. Sci. Fenn. Math. **38** (2013), 3–27.

[10] Tuomas P. Hytönen, *The sharp weighted bound for general Calderón-Zygmund operators*, Ann. of Math. (2) **175** (2012), no. 3, 1473–1506.

[11] ———, *The vector-valued non-homogeneous* Tb *theorem*, Int Math Res Notices, posted on 2012, DOI 10.1093/imrn/rns222, (to appear in print), available at http://www.arxiv.org/abs/0809.3097.

[12] Tuomas Hytönen, *Representation of singular integrals by dyadic operators, and the* A_2 *theorem* (2011), available at http://arxiv.org/pdf/1108.5119.pdf.

[13] Tuomas Hytönen, *A framework for non-homogeneous analysis on metric spaces, and the RBMO space of Tolsa*, Publ. Mat. **54** (2010), no. 2, 485–504.

[14] Tuomas Hytönen, Suile Liu, Dachun Yang, and Dongong Yang, *Boundedness of Calderón-Zygmund operators on non-homogeneous metric measure spaces*, Canad. J. Math. **64** (2012), no. 4, 892–923.

[15] Tuomas Hytönen and Henri Martikainen, *Non-homogeneous* Tb *theorem and random dyadic cubes on metric measure spaces*, J. Geom. Anal. **22** (2012), no. 4, 1071–1107.

[16] ———, *On general local* Tb *theorems*, Trans. Amer. Math. Soc. **364** (2012), no. 9, 4819–4846.

[17] Tuomas Hytönen and Fedor Nazarov, *The local* Tb *theorem with rough test functions* (2012), available at http://arxiv.org/abs/1206.0907.

[18] Tuomas Hytönen and Mark Veraar, R-*boundedness of smooth operator-valued functions*, Integral Equations Operator Theory **63** (2009), no. 3, 373–402.

[19] Tuomas P. Hytönen and Antti V. Vähäkangas, *The local non-homogeneous* Tb *theorem for vector-valued functions* (2012), available at http://arxiv.org/abs/1201.0648.

[20] Michael T. Lacey, Eric T. Sawyer, Chun-Yen Shen, and Ignacio Uriarte-Tuero, *The Two Weight Inequality for Hilbert Transform, Coronas, and Energy Conditions*, available at http://www.arxiv.org/abs/1108.2319.

[21] ———, *Two Weight Inequality for the Hilbert Transform: A Real Variable Characterization*, available at http://www.arxiv.org/abs/1201.4319.

[22] Michael T. Lacey and Antti V. Vähäkangas, *The Perfect Local* Tb *Theorem and Twisted Martingale Transforms*, Proc. Amer. Math. Soc., to appear (2012), available at http://www.arxiv.org/abs/1204.6526.

[23] ———, *On the Local* Tb *Theorem: A Direct Proof under Duality Assumption* (2012), available at http://www.arxiv.org/abs/1209.4161.

[24] Henri Martikainen, *Vector-valued non-homogeneous* Tb *theorem on metric measure spaces*, Rev. Mat. Iberoam. **28** (2012), no. 4, 961–998.

[25] F. Nazarov, S. Treil, and A. Volberg, *The* Tb*-theorem on non-homogeneous spaces*, Acta Math. **190** (2003), no. 2, 151–239.

[26] ———, *Accretive system* Tb*-theorems on nonhomogeneous spaces*, Duke Math. J. **113** (2002), no. 2, 259–312.

[27] Stefanie Petermichl, *Dyadic shifts and a logarithmic estimate for Hankel operators with matrix symbol*, C. R. Acad. Sci. Paris Sér. I Math. **330** (2000), no. 6, 455–460.

[28] Alexander Volberg and Brett D. Wick, *Bergman-type singular integral operators and the characterization of Carleson measures for Besov-Sobolev spaces and the complex ball*, Amer. J. Math. **134** (2012), no. 4, 949–992.

(M.T.L) School of Mathematics, Georgia Institute of Technology, Atlanta GA 30332, USA
E-mail address: lacey@math.gatech.edu

(A.V.V.) Department of Mathematics and Statistics, P.O. Box 68, FI-00014 University of Helsinki, Finland
E-mail address: antti.vahakangas@helsinki.fi